

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**





Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1395 ATTCTTTTGAAGAAATT 1414  
|||||

Db 25015 ATTCTTTTGAAGAAATT 25034

RESULT 45  
AC111865  
LOCUS  
DEFINITION AC111865 101730 bp DNA linear HTG 13-JUL-2002  
Rattus norvegicus clone CH230-214018, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 51 unordered pieces.

ACCESSION AC111865  
VERSION AC111865.2 GI:21736631  
KEYWORDS HTG; HTGS PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 101730)  
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, P.R., Allen, C.,  
Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Ayele, M., Banks, T.,  
Barbaria, J., Benton, J., Bimege, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Deigado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Eamhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Licharge, O., Lieu, C., Lucier, R., Luna, W., Louisse, H.,  
Lozano, R.J., Lu, X., Lucier, A., Martin, R., Martindale, A., Martinez, E.,  
Maheshwari, M., Mapua, P., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Mosser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,  
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, X.,  
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,  
Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tatton, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinsion, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 101730)  
AUTHORS Worley, K.C.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 101730)  
AUTHORS Worley, K.C.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 4 (bases 1 to 101730)  
AUTHORS Worley, K.C.  
TITLE Direct Submission  
JOURNAL Unpublished

of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 12, 2002 this sequence version replaced gi:18701709.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GOJS  
Center clone name: CH230-214018  
----- Summary Statistics  
Sequencing vector: Plasmid,  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 35302 bases at least Q40  
Consensus quality: 38351 bases at least Q30  
Consensus quality: 41540 bases at least Q20

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 51 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1006: contig of 1006 bp in length  
1007: gap of unknown length  
1107: contig of 1020 bp in length  
2127: gap of unknown length  
2227: contig of 1332 bp in length  
3559: gap of unknown length  
3559: contig of 1643 bp in length  
5301: gap of unknown length  
5302: contig of 1028 bp in length  
5402: gap of unknown length  
6430: contig of 1243 bp in length  
6530: gap of unknown length  
7773: contig of 1594 bp in length  
7873: gap of unknown length  
9467: gap of unknown length  
9567: contig of 1278 bp in length  
10844: gap of unknown length  
10945: contig of 1588 bp in length  
12533: gap of unknown length  
12632: contig of 1284 bp in length  
13916: gap of unknown length  
14016: contig of 1286 bp in length  
15302: gap of unknown length  
15402: contig of 1331 bp in length  
16733: gap of unknown length  
16833: contig of 1699 bp in length  
18532: gap of unknown length  
18632: contig of 1374 bp in length  
20006: gap of unknown length  
20106: contig of 1623 bp in length  
21729: gap of unknown length  
21829: contig of 1726 bp in length  
23555: gap of unknown length  
23655: contig of 1298 bp in length  
24953: gap of unknown length  
25053: contig of 1262 bp in length  
26315: gap of unknown length  
26415: contig of 1605 bp in length  
28020: gap of unknown length  
28120: contig of 1338 bp in length  
29458: gap of unknown length  
29558: contig of 1236 bp in length  
30794: gap of unknown length  
30894: contig of 1578 bp in length  
32472: gap of unknown length  
32572: gap of unknown length

Search completed: December 17, 2002, 11:28:26  
Job time : 8694 secs

\* 32573 33967: config of 1395 bp in length  
\* 33968 34067: gap of unknown length  
\* 34068 35428: config of 1361 bp in length  
\* 35429 35528: gap of unknown length  
\* 35529 36887: config of 1359 bp in length  
\* 36888 36987: gap of unknown length  
\* 36988 38865: config of 1878 bp in length  
\* 38866 38966: gap of unknown length  
\* 38967 40597: config of 1632 bp in length  
\* 40598 40697: gap of unknown length  
\* 40698 42321: config of 1624 bp in length  
\* 42322 42421: gap of unknown length  
\* 42422 44368: config of 1947 bp in length  
\* 44369 44468: gap of unknown length  
\* 44469 46059: config of 1591 bp in length  
\* 46060 46159: gap of unknown length  
\* 46160 47945: config of 1786 bp in length  
\* 47946 48045: gap of unknown length  
\* 48046 50298: config of 2253 bp in length  
\* 50299 50398: gap of unknown length  
\* 50399 52618: config of 2220 bp in length  
\* 52719 52718: gap of unknown length  
\* 52719 54903: config of 2185 bp in length  
\* 54904 55003: gap of unknown length  
\* 55004 56112: config of 1109 bp in length  
\* 56113 56212: gap of unknown length  
\* 56213 58045: config of 1833 bp in length  
\* 58046 58145: gap of unknown length  
\* 58146 59642: config of 1497 bp in length  
\* 59643 59742: gap of unknown length  
\* 59743 61011: config of 1269 bp in length  
\* 61012 61111: gap of unknown length  
\* 61112 62667: config of 1556 bp in length  
\* 62668 62767: gap of unknown length  
\* 62768 64883: config of 2116 bp in length  
\* 64884 64983: gap of unknown length  
\* 64984 67363: config of 2380 bp in length  
\* 67364 67463: gap of unknown length  
\* 67464 68975: config of 1512 bp in length  
\* 68976 69075: gap of unknown length  
\* 69076 71513: config of 2438 bp in length  
\* 71514 71613: gap of unknown length  
\* 71614 73900: config of 2287 bp in length  
\* 73901 74000: gap of unknown length  
\* 74001 76769: config of 2769 bp in length  
\* 76770 76869: gap of unknown length  
\* 76870 79077: config of 2208 bp in length  
\* 79078 79177: gap of unknown length  
\* 79178 82287: config of 3110 bp in length  
\* 82288 82387: gap of unknown length  
\* 82388 86144: config of 3757 bp in length  
\* 86145 86244: gap of unknown length  
\* 86245 91344: config of 5100 bp in length  
\* 91345 91444: gap of unknown length  
\* 91446 96049: config of 4605 bp in length  
\* 96050 96149: gap of unknown length  
\* 96150 101730: config of 5581 bp in length.

FEATURES  
source 1..101730  
Location/Qualifiers  
1..101730  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-214018"

BASE COUNT 29228 a 18778 c 18804 g 29744 t 5176 others

Query Match 1.1%: Score 20; DB 2; Length 101730;  
Best Local Similarity 100.0%: Pred. No. 38;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1650 ATTTGCAATTATATGACC 1669  
|||||  
DB 30425 ATTTGCAATTATATGACC 30444

Query Match: 86.78% Indels: 3  
DB: 2 Gaps: 0  
US-09-868-987-1 (1-1864) x E72062 (1-1402)

QY 2 TGGACATTCGCGCATATGTCAGAGGGGATCAGCGCACTGCTAATGACATATTCGCA 61  
DB 817 TrpThrSerAlaTyrCysGlnGluGlyIleSerGlyThrAlaAsnGlyGlnTyrSerAla 836

QY 62 AACCGTGGATGGCGTATGCGTGTAGTCAATGATTCAGCGGTATATGTCAGCAGCGCCCTATTTTA 121  
DB 837 AsnArgGlyTrpArgMetAlaValIleAspGlyTyrMetValSerSerProIleLeu 856

QY 122 AACGTCCCATTTGAAATCATCGCAGTGTCTCAGGGAATTTTACCACCGTGAAGTGAAGC 181  
DB 857 AsnValProLeuLysAsnHisAlaSerValSerGlyLysPheThrHisArgGluValSer 876

QY 182 AAATCGCGCTCAGATTTAAATCTGGAGCGATGCTTTTGTTCGCGAGGTCTCAGTGA 241  
DB 877 LysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheValProGluValLeuSerGlu 896

QY 242 GAGACGATCTCTCTGATCTTTGGGAAAAAACAATGTACACAAGGCATTTATCTCAGCATGC 301  
DB 897 GluThrIleSerSerAspLeuGlyLysGlnCysThrGlnGlyIleIleSerAlaCys 916

QY 302 TGTGCGTGGCAATGCTTATGTTTGTAGAGCGTATATATAGATTTGGAGGCGTCAATC 361  
DB 917 CysGlyLeuAlaMetLeuIleValLeuMetSerValTyrTyrArgPheGlyGlyValIle 936

QY 362 GCTTCGGGAGCTGCTCTCAATCTTTGCTTATCTGGCGAGCTCTACAGTATTCGAT 421  
DB 937 AlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrpAlaLeuLeuGlnTyrLeuAsp 956

QY 422 GCGCACTCACCTTGTGAGGACTCGCTGGGATGTTCTTGTATGGGATGGCGGTAGAT 481  
DB 957 AlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaValAsp 976

QY 482 GCAATGTTCTGTATTCGAAAGAAATCCGAGAGAAATTTTATGTTCTCAAAGTCTTAAA 541  
DB 977 AlaAsnValLeuValPheGluArgIleArgGluGluPheLeuLeuSerGlnSerLeuLys 996

QY 542 AAATCTGTAGAAAAGGATATACCAAGCTTTTCGAGCCATTTTGTATCTCACTTGAAT 601  
DB 997 LysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeuThr 1016

QY 602 ACAGTATTTGGCTCAGCACTTCTTCTCTCTAGATACAGGCGCTATTAAAGGGTTTGCT 661  
DB 1017 ThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPheAla 1036

QY 662 TTGACATTTAGGAATTTCTCTTCAATGTTTACGGCTCTTTTCATGACTTAAATTT 721  
DB 1037 LeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLysPhe 1056

QY 722 TTCTTCATGCTGTCGATGAATTAAGACCAACATACACAGTTGCTATGATGAATAGTTTC 781  
DB 1057 PhePheMetLeuTrpMetAsnLysThrGlnHisThrGlnLeuHisMetMetAsnLysPhe 1076

QY 782 GTGGGGATAAAGCATGATTTCTTCAGAGAGATGCAAAAACCTTCGGCTGTTCTTCGGAAT 841  
DB 1077 ValGlyIleLysHisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGlySer 1096

QY 842 GTTTTCTTTTAGTGTGCTCTCGGTTTCGAGCTCGGAATTCGGTTTTCGGGAATG 901  
DB 1097 ValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMet 1116

QY 902 GATTTTAAAGAGGGTATGCTTTACCTTTTAAATCCAAAAGCATGGCATCAGGATGTT 961  
DB 1117 AspPheLysGlyTyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspVal 1136

QY 962 GCTCAATGCGGCAAGTGTGCATAACTACAGGAAGCTGCTTCTTCTCTAGAGAC 1021  
DB 1137 AlaGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeuSerSerArgAsp 1156

QY 1022 TTCGCTATTCAAACATTTGGATCTTCAGAAAAGATCAAATCTATTTTAGTGATAAGCT 1081  
DB 1157 PheArgIleGlnThrPheGlySerSerGluLysIleLysIleTyrPheSerAspLysAla 1176

QY 1082 TTAAGCTATATAAGCA-GATACGAGCTCTCTCCTAAAAATTAAACGATCATGAGCTGGCG 1140  
DB 1177 LeuSerTyrThrLysAlaAspThrSerLeuSerProLysIleAsnAspHisGluLeuAla 1196

QY 1141 TTATT-GTGGATTGTTGTCAGAAACAGCGCTAGATTCTCTACGGAAACTCTAAACGAA 1199  
DB 1197 LeuAlaValGlyLeuLeuSerGluThrGlyLeuAspPheSerThrGluThrLeuAsnGlu 1216

QY 1200 ACCGAAAA-TTTTGGTCAAGGTAAGCAGCAAACTATCGAAGAAAATGCGTTATCATGCGC 1258  
DB 1217 ThrGlnAsnPheTrpSerLysValSerSerLysLeuSerLysLysMetArgTyrGlnAla 1236

QY 1259 ACCATCGGGCTTTTAGGAGCTTTGGCAATCATCTTGCTCTATGTGAGTTTGGCGTTTGA 1318  
DB 1237 ThrIleGlyLeuLeuGlyAlaLeuAlaIleIleLeuLeuTyrValSerLeuArgPheGlu 1256

QY 1319 TGGCAATATGCTTTCACTGCGCTGATGCGCTTTAATTCATGACCTTTTGGCTACCTGTGCA 1378  
DB 1257 TrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAla 1276

QY 1379 GTCTGTTTATAGCACATTTCTTTTGAAGAAAATTCAAATAGATTTGCAAGCCATTTGGT 1438  
DB 1277 ValLeuPheIleAlaHisPhePheLeuLysIleGlnIleAspLeuGlnAlaIleGly 1296

QY 1439 GCTTTAATGACTGTTATGGGGTATTCATTAACAATACTTTGATCATTTTTGTGATCGATT 1498  
DB 1297 AlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIle 1316

QY 1499 CGTCAAGATCCCAAGCAACCTGTTTACCCCTATGATGCTTTTAGTTAATGATGCCCTT 1558  
DB 1317 ArgGluAspArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeu 1336

QY 1559 CAAAAGAGTTTCAGCCCGCCAGGTAAAGAACAGCTACAACTCTATCAGTTTGTAAATG 1618  
DB 1337 GlnLysThrPheSerArgThrValMetThrAlaThrThrLeuSerValLeuLeuMet 1356

QY 1619 CTTTGTGTTTATAGCGGCTCTCTGCTTTAATTTTGCATTTATTATGACCATAGGATTT 1678  
DB 1357 LeuLeuPheIleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIle 1376

QY 1679 CTTCTAGGAACCTTATCGTCTCTTTTATATTCACCACTCTGTTGTTTATGTCGTCGT 1738  
DB 1377 LeuLeuGlyThrLeuSerSerLeuTyrIleAlaProLeuLeuLeuPheMetValArg 1396

QY 1739 AAAGAAAATCGCTCAAAA 1756  
DB 1397 LysGluAsnArgSerLys 1402

RESULT 3  
A71514  
probable protein export - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C;Species: Chlamydia trachomatis  
C;Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
C;Accession: A71514  
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998  
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis  
A;Reference number: A71570; MUID:99000809; PMID:9784136  
A;Accession: A71514  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1400 <ARN>  
A;Cross-references: GB:AB001318; GB:AB001273; NID:G3328875; PIDN:AA68047.1; PID:G3328880  
A;Experimental source: serotype D, strain UW-3/Cx  
C;Genetics:  
A;Gene: secD/secF  
Alignment Scores: 1.73e-174 Length: 1400  
Pred. No.:

Score: 2083.00 Matches: 415  
Percent Similarity: 79.28% Conservative: 67  
Best Local Similarity: 68.26% Mismatches: 80  
Query Match: 62.89% Indels: 46  
DB: 2 Gaps: 5

US-09-868-987-1 (1-1864) x A71514 (1-1400)

QY 2 TGAAGCTCCGATATTTGACAGAGGGGATCACCGGCACTGCTAAATGACAAATATTTGCA 61  
DB TrpThrSerLysPheCysGlnGluValSerGlyThrLysAsnGlyLeuPheSerLys 831  
QY 62 AACGGTATGATGCGATGCTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 121  
DB GlyArgGlyTrpArgMetAlaValAlaLeuAspGlyTrpValIleSerAspProValLeu 851  
QY 122 AACGTCATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAA 181  
DB AsnValProLeuLysAspHisAlaSerValSerGlyAsnPheSerTrpArgGluValHis 871  
QY 182 AAATCGCTCAGATTTAAATCTGAGAGCATGCTTTTGTCCGAGCTTCTCAGTGA 241  
DB ArgLeuAlaAlaAspLeuLysSerGlyAlaMetSerPheIleProGluIleLeuSerGlu 891  
QY 242 GAGACGATCTCTGATCTTGGGAAAAACAATGATACACAGCATATCTCAGCATGC 301  
DB GluValValSerProGluLeuGlySerSerGlnArgValGlnGlyLeuSerValVal 911  
QY 302 TGTGGCTTGGCAATGCTTATTTGATGAGCGATATATATATATATATATATATATATATAT 361  
DB LeuGlyLeuValValLeuIleValLeuMetSerValTrpTrpArgPheGlyGlyValIle 931  
QY 362 GCTTCGGAGCTGTTCTTGTGAATCTTTGCTTATCTGGGCACTCTCAGATTTGAT 421  
DB AlaSerIleAlaValLeuLeuAsnLeuLeuLeuIleTrpAlaSerMetGlnTrpLeuAsp 951  
QY 422 GGGCAGCTCAGCTTGTGAGGACGCTGGGATGTTCTTGTGATGAGGAGGAGGAGGAGGAG 481  
DB AlaProLeuThrLeuSerGlyLeuAlaGlyIleIleLeuAlaMetGlyMetAlaValAsp 971  
QY 482 GCAATGCTTCTTGTATGAAAGAAATCCGAGAGAAATTTTATGCTCAAAAGTCTTAA 541  
DB AlaAsnValLeuValPheGluArgGlyLeuArgGluGlyTrpLeuLeuThrArgSerLeuSer 991  
QY 542 AAATCTGAGAAAAAGATATACCAAGGCTTTGAGGCACTTTGATTTGATCTTACTGACT 601  
DB GluSerValGluAlaGlyTrpLysLysAlaPheSerAlaIlePheAspSerAsnLeuThr 1011  
QY 602 ACAGATTTGGCTCGAGCACTTCTTCTTCTCAGATACAGGCGCTATTAAGGGTGTGCT 661  
DB ThrIleLeuAlaSerAlaLeuLeuLeuMetLeuAspTrpGlyProIleLysGlyPheAla 1031  
QY 662 TTGACATGATTTTGAAGATTTTCTTCAATGTTTACGGCTCTTTGATGACTAAATTT 721  
DB LeuThrIleuIleIleGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLysPhe 1051  
QY 722 TTCTTCATGCTGTGATGATATAGACCAACATACACAGTTGCATATGATGATATAGTTC 781  
DB PhePheValIleTrpValGlnLysThrArgGluTrpGlnLeuHisMetLeuAsnLysPhe 1071  
QY 782 GTGGGATTAAGCATGATTTCTTGAAGAGATGCAAAAACCTTGGGCTGTTTCTGAGAGT 841  
DB IleGlyIleLysHisAsnPheLeuLysGluCysLysAspGlyLeuTrpValValSerIleVal 1091  
QY 842 GTTTTCTTTTGAAGTGTGCTCTCGGATTTGAGGCTGGAATCCGTTTGGGAGT 901  
DB ValLeuValIleGlyCysValGlyLeuGlyPheGlyAlaTrpAspSerValPheGlyMet 1111  
QY 902 GATTTTAAGAGGGGTATGCTTTTATCTTAAATCCAAAAGCATGATGATGATGATGATGAT 961  
DB AspPheLysGlyGlyTrpAlaLeuThrLeuAsp-----SerAspVal 1125  
QY 962 GCT-----CAATGCGTGGCAAAAGTTGTGATAACTACAGGAAGCTGCT 1006

DB 1126 CysGluTrpAsnProGluGlnMetCysSerValLeuArgLysArgPheGlnGlnIleGly 1145  
QY 1007 CTTTCTTCTGAGACCTTCGATATTAACAACTTTGATCTTCAGAAAAATCAAAATCTAT 1066  
DB LeuSerSerArgAspTrpArgValArgArgAlaAspSerSerGluLysValLysIleTrp 1165  
QY 1067 TTTTATGATTAAGCTTAACTATCTAAGCAGATACGA----- 1105  
DB LeuSerGlnAsnAlaLeuAspArgValGluGlnIleGluGlyAlaGlySerGluGlnLys 1185  
QY 1106 -----GCTCTCTCTTAAATTAACATGATGAGCTGGCTTATTTGT 1147  
DB GlySerAspTrpHisLeuAlaArgValLeuGlnValLeuSerAspSer----- 1201  
QY 1148 GGGATTTGTTGTCAGAAAACAGGCTCAGATTTTCTCAGGAAACTCTTAAACGAAACGAAA 1207  
DB 1202 -----GlySerSerThrThrSerMetVal 1209  
QY 1208 TTT-----TGTCAAGGTAAGCAAGCAAACTATGAGAAAGAAATGCGT 1249  
DB PheAspAlaSerArgGlySerTrpPheLysValSerIleGlnLeuSerAsnLysMetArg 1229  
QY 1250 TATCAGCGCAGCAATCGGGCTTTTGAAGCTTTGCAATCATCTTGTCTATGATGATTTG 1309  
DB ThrGlnAlaValIleAlaLeuPheGlyAlaLeuGlyIleIleLeuPheTrpValSerLeu 1249  
QY 1310 CGCTTGAATGCAATATGCTTTCAGATGCGGATGCGCTTTAATCATGACCTTTGGCT 1369  
DB ArgPheGluTrpArgTrpArgAlaPheSerAlaIleCysAlaLeuMetHisAspLeuValAla 1269  
QY 1370 ACCGTGACAGCTTGTATATAGCACATTTCTTTTGAAGAAATTCAAATATGATTTGCA 1429  
DB ThrCysAlaValLeuValAlaLeuHisPhePheLeuGlnArgIleGlnIleAspLeuGln 1289  
QY 1430 GCCATGAGCTTATATGACTGATTTGAGGATTTTATTAACAATATCTTATGATTTT 1489  
DB AlaIleGlyAlaLeuMetThrValLeuGlyTrpSerLeuAsnHisThrIleIlePhe 1309  
QY 1490 GATGCTATTTGTTGAAGATCGGCAAGCAAGCTGTTTATCCCTATGATGATTTTATGTTAAT 1549  
DB AspArgIleArgGluAspArgValArgGlyLeuLeuPheThrProMetProIleLeuIleAsn 1329  
QY 1550 GATGCCCTTCAAAAAGAGCTTCAAGCGGAGGATATGACACAGCTATACCTATACGTT 1609  
DB AspAlaLeuGlnLysThrLeuGlyArgTrpValMetThrThrAlaThrThrLeuSerVal 1349  
QY 1610 TTGTTAATGCTTGTATATAGGCGGCTCCTGCTTATATTTTGCATTTATATGACC 1669  
DB LeuValIleLeuLeuPheValGlyGlyGlySerIlePheAsnPheAlaPheIleMetThr 1369  
QY 1670 ATAGGAATCTTCTGAGAACCTTATATGCTCTTATATATGACACCACTCTGTTGTTT 1729  
DB ValGlyIleLeuLeuGlyTrpLeuSerSerLeuTrpIleAlaProLeuLeuPhe 1389  
QY 1730 ATGTCGCTAAAGAAATCGCTCA 1753  
DB 1390 MetValArgLysGluGlnAsn 1397

RESULT 4  
A81672  
seclP protein, probable TCO733 [imported] - Chlamydia muridarum (strain Nigg)  
C/Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C/Date: 31-Mar-2000 #sequence\_rev150n 31-Mar-2000 #text\_change 11-May-2000  
C/Accession: A81672  
R/Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwynn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg  
Nucleic Acids Res. 28, 1397-1406, 2000  
A/Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A/Reference number: A81672  
A/Accession: A81672  
A/Status: preliminary  
A/Molecule type: DNA

A;Residues: 1-1400 &lt;TET&gt;

A;Cross-references: GB:AB002341; GB:AB002160; NID:97190754; PIDN:AAF39543.1; PID:g719076

A;Experimental source: strain Nigg (MoPn)

C;Genetics:

A;Gene: TC0733

## Alignment Scores:

Pred. No.:	3,58e-173	Length:	1400
Score:	2068.00	Matches:	407
Percent Similarity:	79.61%	Conservative:	77
Best Local Similarity:	66.94%	Mismatches:	78
Query Match:	62.44%	Indels:	46
DB:	2	Gaps:	5

US-09-868-987-1 (1-1864) x AB1672 (1-1400)

```
QY 2 TGGACTCCGCATATTGTTCAGAGGGGATCAGCGCACTGCTAATGGACAATATTCTGCA 61
DB 812 TrpThrSerLysPheCysGlnGluGlyValAsnGlyThrLysAsnSerGlnPheSerGly 831
QY 62 AACCGTGGATCGCGTATGGCTGTAGTGTACGGTATATGGTTCAGCAGCGCCTATTTTA 121
DB 832 GlyArgGlyTyrArgMetAlaValValLeuAspGlyTyrValIleSerAspProValLeu 851
QY 122 AACGTCCCAATTGAAAATCATGCCAGTGTCTCAGGGAATTTTACCACCGTGAAGTGAGC 181
DB 852 AsnValProLeuLysAspHisAlaSerValSerGlyAsnPheSerTyrArgGluValAsn 871
QY 182 AAACCTCGCTCAGATTTAAATCTGGACGATGCTTTTGTTCACGAGGTTCTCAGTGAA 241
DB 872 ArgLeuAlaSerAspLeuSerGlyAlaMetSerPheIleProGluIleLeuSerGlu 891
QY 242 GAGACGATCTCTCTGATCTTGGGAAAAAACAATGTACACAGGCAATTTCTCAGCATGC 301
DB 892 GluValValSerProGluLeuGlyHisSerGlnArgLeuGlnGlyLeuSerValPhe 911
QY 302 TGTGGCTGGCAATGCTTATTGTTTGATAGCGGTATATTATAGATTGGAGCGGTATC 361
DB 912 LeuGlyLeuValValLeuIleAlaLeuMetSerValTyrLysPheGlyGlyValIle 931
QY 362 GCTTCGGGAGCTGTTCTCTGAATCTTTGCTATCTTCGGGAGCTCTACAGTATTGGAT 421
DB 932 AlaSerValAlaValLeuLeuAsnLeuLeuIleTyrPheAlaSerMetGlnTyrLeuAsp 951
QY 422 CGGCACCTCACCCTGTTCAGGACTCGCTGGGATGTTCTTGTGATGGGATGGCGGTAGAT 481
DB 952 AlaProLeuThrLeuSerGlyLeuAlaGlyIleIleLeuAlaMetGlyMetAlaValAsp 971
QY 482 GCAATGTTCTTGATTCGAAAGAAATCCGAGAGGAATTTTATTGTCACAAAGCTTAAA 541
DB 972 AlaAsnValLeuValPheGluArgIleArgGluGluTyrLeuLeuSerArgSerLeuSer 991
QY 542 AAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGGCATTTTGTATCTTAACTTGACT 601
DB 992 GluSerValGluAlaGlyTyrLysAlaPheSerAlaIlePheAspSerAsnLeuThr 1011
QY 602 ACAGTATTGGCTCAGCACTCTCTTCTCTAGATACAGGCGCTTATAAAGGGTTTCT 661
DB 1012 ThrIleLeuAlaSerAlaLeuLeuLeuMetLeuAspThrGlyProIleLysGlyPheAla 1031
QY 662 TTGACATGATTTTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATGACTAAATTT 721
DB 1032 LeuThrLeuIleIleGlyIlePheSerSerMetPheThrSerLeuPheMetThrLysPhe 1051
QY 722 TCTCTCATGCTGTGATGAATAAGCCCAACATACACAGTTGCATATGATGAATAGTTC 781
DB 1052 PhePheLeuIleIleTyrIleGlnLysThrArgGluThrGlnLeuHisMetMetAsnLysPhe 1071
QY 782 GTGGGGATAAAGCATGATTTCTTGTAGAGGATGCAAAAAAATCTTGGGCTGTTCTGGAGGT 841
DB 1072 IleGlyValLysHisAsnPheLeuLysGluCysLysArgLeuTyrValValSerGlyAla 1091
QY 842 GTTTTCTTTTAGGTGCTGCTCTCGGGTTTGGAGCCTCGGAATTCGGTATTGCGGAATG 901
```

```
DB 1092 ValValValLeuGlyCysIleGlyLeuGlyPheGlyAlaTrpAspSerIleLeuGlyMet 1111
QY 902 GATTTTAAAGAGGAGGTATGCTTTTACCTTTAAATCCAAAAGAGCATGCGCATCAGCGATGTT 961
DB 1112 AspPheLysGlyGlyTyrAlaLeuThrLeuAsp-----SerAspVal 1125
QY 962 GCT-----CAAATGCGTGGCAAAAGTTGTGCATAAACTACAGGAAGCTGGT 1006
DB 1126 CysAlaTyrAsnProAspGlnMetCysSerAlaLeuLysValysArgPheArgGluIleGly 1145
QY 1007 CTTTCTTCTAGACACTTCCGTATTCAACATTTGGATCTTCAGAAAAGATCAAAATCTAT 1066
DB 1146 LeuSerSerArgAspTyrArgValArgLysAlaAspSerSerGlyLysValLysIleTyr 1165
QY 1067 TTTAGTCATAAAGCTTTTAAAGCTATATAAGCAGATACGA----- 1105
DB 1166 PheSerGlnAsnAlaLeuAlaArgValGluGlnAlaLysGlyTyrSerGluGluAsn 1185
QY 1106 -----GCCTCTCTCTAAATTAACGATCATGAGCTGGCGTATTGTT 1147
DB 1186 GlyAlaAspHisHisLeuAlaGlnValLeuGlnValLeuSerAspSer----- 1201
QY 1148 GGGATTCTTGTCAAAAACAGCGCTAGATTCTCTACGGAACCTCTAAACGAAACGCAAAA 1207
DB 1202 -----SerAspAlaThrSerProAlaAla 1209
QY 1208 TTT-----TGGTCAAAAGTAAAGCAGCAAACTATCGAAGAAAATGCGT 1249
DB 1210 PheGluSerSerArgGlySerTrpPheLysValSerGlyGlnLeuSerAsnLysMetArg 1229
QY 1250 TATCAGCGCAGCAATGCTTTTAGGAGCTTTTGGCAATCATCTTTCGCTCTATGTGAGTTG 1309
DB 1230 ThrGlnAlaIleMetAlaLeuPheGlyAlaLeuGlyIleIleLeuLeuTyrValSerLeu 1249
QY 1310 CGCTTGAATGGCAATATGCTTTTTCAGTCCGCTATGCGCTTTAATTCATGACCTTTTGCT 1369
DB 1250 ArgPheGluTyrArgTyrAlaPheSerAlaIleCysAlaLeuMetHisAspLeuLeuAla 1269
QY 1370 ACCTGTGCGAGCTTTTATAGCACATCTCTTTTGAAGAAATTCAAATAGATTGCAA 1429
DB 1270 ThrCysSerValLeuValAlaLeuHisPhePheLeuHisArgIleGlnIleAspLeuGln 1289
QY 1430 GCCATTGCTGCTTTAATCAGCTATGCGGCTATTTCATTAACAATACCTTTGATCATTTT 1489
DB 1290 AlaValGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePhe 1309
QY 1490 GATCGTATTGCTGAAGATCGCCAAAGCGAAGCTGTTTACCCTATGCTGTTTATAGTTAAT 1549
DB 1310 AspArgIleArgGluAspArgArgGluLysLeuPheThrProMetProIleLeuIleAsn 1329
QY 1550 GATGCCCTTCAAAAGAGCTTCAGCGCAGCGTATGACACAGCTACACTCTATCAGTT 1609
DB 1330 AspAlaLeuGlnLysThrLeuGlyArgThrValMetThrThrAlaThrLeuSerVal 1349
QY 1610 TTGTTAATGCTTTTGTATTATAGCGGCTCTCTCTCTTTAATTTTGCATTTTATATGACC 1669
DB 1350 LeuValIleLeuLeuPheValGlyGlySerIlePheAsnPheAlaPheIleMetThr 1369
QY 1670 ATAGGGATCTTCTAGGAATTTTATCGTCTCTTTTATATTGCACCCTCTCTGTTGTTT 1729
DB 1370 ValGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeuPhe 1389
QY 1730 ATGGTCCGTAAGAAATCGCTCA 1753
DB 1390 MetValArgLysGluGluGlnAsn 1397
RESULT 5
AH2768
protein-export membrane protein SECD [imported] - Agrobacterium tumefaciens (strain C58,
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AH2768
```

R:Wood, D.W.; Serubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; erge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCellan, Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.

A>Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AH2768

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1,857 <KUR>

A:Cross-references: GB:AB008688; PIDN:AL42566.1; PID:G17739990; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: secD

A:Map position: circular chromosome

Alignment Scores:

Pred. No.:	2,21e-65	Length:	857
Score:	840.00	Matches:	198
Percent Similarity:	57.75%	Conservative:	141
Best Local Similarity:	33.73%	Mismatches:	209
Query Match:	25.36%	Indels:	39
DB:	2	Gaps:	13

US-09-868-987-1 (1-1864) x AH2768 (1-857)

```

QY 7 TTCCGCATATTGTCAGAGGAGATCGCGCATGCTGTAATGACAAATATTCTGCAACCG 66
Db 282 PheargPheaspserargylalaglnarthe-Alaglnlarthe---GlnGlnaserva 300
QY 67 TGGATGGCGTATGCTGATGATGACGCTTATATGCTACAGCCCTATTAAACGT 126
Db 300 lGlylserprohealalvalleuaspsnglnvalilleseeralaprovalilleargl 320
QY 127 CCCA---TTGAAAATCATGCCAGTCTCGAGGAATTTCGCCAGCTGAAGAGCAA 183
Db 320 uprilellellylserglylnlleserllyserpheservalGlnGlyalalaspas 340
QY 184 ACTGCGCTCAATTTAAATCTGAGCGATGCTTTTGTCCCGAGCTTCAGTGAAGA 243
Db 340 pleuhlaValleuenaalaglyalaleuProalartheuthrValalagluuar 360
QY 244 GACGATCTCTTCTGATCTTGGGAAAAAACAATGACACAGGCTTATTCACATGCTG 303
Db 360 gThrValGlyProserleuGlyAsnaspserillethralaglyleuthralaseralail 380
QY 304 TGGCTGGCATGCTTATGTTGATGACGCTATATATGATTTGGAGCGCTCATGCG 363
Db 380 eGlyalalalaglyalalleuilepemetPhevalPheValGlyPhePheGlyleuual 400
QY 364 TTCCGAGAGCTCTTCTGCAATCTTTGCTTATCTGCGACCTCATACGATTTGATGC 423
Db 400 aasnillealaleuilevalasnaValalMetleuilealaleuaserValilleglyse 420
QY 424 GCCACTCACCTTGACGAGCTCGCTGGAGATTGTTCTGTAAGGGAGTGGCGTATGTC 483
Db 420 rThleuthrleuproglyllealaglyllealValleuthrilleglyMetalaValaspse 440
QY 484 AAAATGTTTGTATTCGAAAGAAATCCGAGAGGAATTTTATGTCTCAAGCTTAAAAA 543
Db 440 rAsnValleuiletyrGlyargylearGlyluvalylserserlylProleuilegl 460
QY 544 ATCTGTGAAAAAGATATACAGGCTTTTGGAGCCATTTTGTATTTGATTTGATGAC 603
Db 460 nAlleuaspsnglylPheThrargalalPhealalThrilleleaspaalaleuThTh 480
QY 604 AGATTTGGCTCAGACATCTTTTCTCTGATATACAGGCGCTTATTAAGGGTTGCTTT 663
Db 480 rleuileValalaserValleuPheTymerGlyThrGlyProvallylsglyPhealalava 500
QY 664 GACATGATTTTAGAATTTTCTCTCAATGTTTACGGCTTTTCAATGATTAATTTT 723

```

```

Db 500 lThrleualalalaglyllellethrThrValPheThrAlaTyThrleuthrAlaTrpme 520
QY 724 CTTCAGCTGTGATGATTAAGACCAACATACAGTTGCAT-----ATGATGAA 774
Db 520 tPheGlylTyThrValargserarGProlySH:leuProlyGlyValargThrAl 540
QY 775 TAAGTTCGTGGGATTAAGATGATTTCTTGAGAGATGCAAAAACTTGGGCTGTC 834
Db 540 ametPheaspGlyArgaspilleProPheMetalarGlyrArgarValalPheMetilleth 560
QY 835 TGGAGTGTCTTTCTTTAGGTTGCGTGTCTCGGGTTGGAGCCCTGGAATTCGGTTT 894
Db 560 rGly---ValilleMetleualacyValagly---GlyPhevalAlalyselyleuasnle 578
QY 895 GGGATGATTTTAAAGAGGATATGCTTACCTTAACTCAAAAGAGCATGCGATCAG 954
Db 578 uGlylleaspPheGlnGlylYlserValilleGluValargAlalagln---GlyGluAl 597
QY 955 CGATGTGCTCAATGCTGTGCGCAAGTTGTCATTAACATACAGGAAGCTGCTTCTTC 1014
Db 597 aaspLeualalaspilleargy-----GluArgleuasnglnleuasnleugly-- 612
QY 1015 TAGAGACTTCGCTATTCAAACATTTGGATCTTCAGAAAAGATCAAAATCTATTAGTGA 1074
Db 613 ---GluilleGlnalaglnasPheGlyserProGlnaspaValleulleargyle---- 629
QY 1075 TAAAGCTTTAAGCTATATACAGATACAGAGCTCTCTCTTAATTAACATCATGAG 1134
Db 630 -GlnAlalaglnaspGlylYlYlGlnasnaalaglnGlnserAlallethreun----- 645
QY 1135 CTGGCGTATTTGGGATTTGCTCGAAGAACAGGCTTCAAGTTTCTCTACGGAACCTTA 1194
Db 646 -----ValArglyGlyleu-----Gluaspy 653
QY 1195 ACGAAGCAAAATTTTGTCAAGGTAGACAGCAAACTATCGAAGAAATGCGTTATCA 1254
Db 653 rTyrasPheargarValalGluValalGlyProalalalaserGlyaspLeuthrValm 673
QY 1255 GCGCAGCATCGGCTTTAGAGCTTTGGCAATCATCTTGTATGTGATGCTTGGCGCT 1314
Db 673 rserThrilleGlyValleuenaalamealalalilleMetvallyrillethrValargph 693
QY 1315 TGAATGCAATATGCTTTCAGTGCCTGATGCGCTTATATCAAGCTTTTGGCTACTG 1374
Db 693 eGluThrGlnPhealaleuGlyalalalleSerMetValHisaspValalPheThr-- 712
QY 1375 TGCAGTCTGTTATATGACATTTCTTTTGAAGAAATTCAAATATGATTTGCAAGCAT 1434
Db 713 -----illeGlyleupheValPheleuglylleGluPheasneuthrseril 728
QY 1435 TGGCTTTAATGATGCTATTTGGGATTTGATTAACATATCTTATGATCATTTTGAATG 1494
Db 728 ealalalleuthrillethleGlyrYrserleuasnaaspThrValalilletyrApar 748
QY 1495 TATTCGTAAGATGCGCAAGCAACCTGTTACCCCTATGACATGTTTACTATATGATGC 1554
Db 748 gValArgGlnasleuargarGlyrlylsyewetProleuSerMetlelleaspValse 768
QY 1555 CTTCAAAAGAGCTTCAGCGCAAGGTAATGACAAACAGTACAACTATACAGTTTGTGT 1614
Db 768 rleuasnglnThrleuserargThrilleleuthrGlyleuthrValleuenaalaleu 788
QY 1615 AATGCTTTGTTATATAGCGCGCTCTCTGTTTAAATTTGATTTATATTAAGACCATAG 1674
Db 788 uAlaleuthrleupheGlylYlGluValilleargserPheThrPhealaleuPhegl 808
QY 1675 GATCTCTTGAAGACTTATGCTCTCTTATATGACCAACCTCTGTGTG---TTAT 1731
Db 808 yValGlylleGlyValPheaserValrlyllealalalaprovalleuilealPhe 828
QY 1732 GGTCCGTAAGAAATCGC 1750

```

Db 828 sLeuArgProaspSerLys 834

## RESULT 6

B97549  
probable protein-export membrane protein [imported] - Agrobacterium tumefaciens (strain C; Species: Agrobacterium tumefaciens  
C; Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C; Accession: B97549  
R; Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A; Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
A; Reference number: A97359; PMID:11743194  
A; Accession: B97549  
A; Status: preliminary  
A; Molecule type: DNA  
A; Residues: 1-876 <KUR>  
A; Cross-references: GB:AE007869; PIDN:AAK87347.1; PID:g15156649; GSPDB:GN00169  
C; Genetics:  
A; Gene: AGR\_C\_2877  
A; Map position: circular chromosome

## Alignment Scores:

Pred. No.:	2, 22e-65	Length:	876
Score:	840.00	Matches:	198
Percent Similarity:	57.75%	Conservative:	141
Best Local Similarity:	33.73%	Mismatches:	209
Query Match:	25.36%	Indels:	39
DB:	2	Gaps:	13

US-09-868-987-1 (1-1864) x B97549 (1-876)

Qy	7	TTCCGCATATTGTTCAGGAGGGGATCAGCGGCACTGCTAATGCAATATTCTGCAAAACCG	66
Db	301	PheArgPheAspSerArgGlyAlaGlnArgPhe-AlaGlnAlaThr---GlnGlnAsnVa	319
Qy	67	TGGATGGCGTATGCTGTAGTGTATGAGTGTATGCTAGCGCCCTATTAAACGT	126
Db	319	IGlyLysProPheAlaIleValLeuAsnGlnValIleSerAlaProValIleArgG	339
Qy	127	CCCA---TTGAAAATCATCGCAGTGTCTCAGGGAATTTTACCACCGTGAAGTCAGCA	183
Db	339	uProIleIleGlySerGlyGlnIleSerGlySerPheSerValGlnGlyAlaAsnAs	359
Qy	184	ACTCGCCTCAGATTAAATCTGGAGCGATGCTTTTGTCCCGAGGTTCTCAGTGAAGA	243
Db	359	pLeuAlaValLeuLeuArgAlaGlyAlaLeuProAlaThrLeuThrValValGluAr	379
Qy	244	GACGATCTCTCTGATCTGGGAAAAACAATGTACACAAGGCATTATCTCAGCATGCTG	303
Db	379	gThrValGlyProSerLeuGlyAsnAspSerIleThrAlaGlyLeuThrAlaSerAla	399
Qy	304	TGGCTTGGCAATGCTATTGTTTGTATGAGCGGTATATATAGATTGGAGCGTCATCGC	363
Db	399	eGlyAlaValGlyValLeuIlePheMetPheValPheThrGlyPhePheGlyLeuLeuAl	419
Qy	364	TTCCGGAGCTGTTCTCTGAATCTTTTCTGCTATCTGGGCGCTCTACAGTATTGGATGC	423
Db	419	aAsnIleAlaLeuIleValAsnValValMetLeuIleAlaValLeuSerValIleGlySe	439
Qy	424	GCCACTACCTTGTGACGACTCGCTGGGATGTTCTTGTCTATGGGATGGCGGTAGATGC	483
Db	439	rThrLeuThrLeuProGlyIleAlaGlyIleValLeuThrIleGlyMetAlaValAepSe	459
Qy	484	AAATGTTCTCTATTCGAAGAATCCGAGAGGAATTTTATGTCTCAAGCTCTTAAAA	543
Db	459	rAsnValLeuIleThrGluArgGileArgGluGluValLysSerGlyLysProLeuIleG	479
Qy	544	ATCTGTAGAAAAAGGATATACCAAGCTTTTGGAGCCATTTTGTATTCTAATCTGACTAC	603
Db	479	naLeuAsnAspGlyPheThrArgAlaPheAlaThrIleIleAsnLeuAsnLeuThr	499
Qy	604	AGTATTGGCCTCAGCACTCTTTTCTCTCTAGATACAGGCGCTATTAAAGGGTTGCTTT	663

Db	499	rLeuIleValAlaSerValLeuPheThrMetGlyThrGlyProValLysGlyPheAlaVa	519
Qy	664	GACATGATTATTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATGACATAATTTT	723
Db	519	lThrLeuAlaValGlyIleIleThrValPheThrAlaThrLeuThrAlaTrpMe	539
Qy	724	CTTCATCTCTGGATGAATTAAGACCCCAACATACACATTTGCAT-----ATGATGAA	774
Db	539	trPheGlyTrpValArgArgSerArgProLysHisLeuProLysGlyValArgThrAl	559
Qy	775	TAAGTTCTGGGATTAAGCATGATTTCTTGAGAGGATGCAAAAACTTTGGGCTGTTTC	834
Db	559	aMetPheAspGlyArgAspIleProPheMetArgTrpArgValValPheMetIleTh	579
Qy	835	TGGAAGTGTCTTTTCTTTAGGTTGCTCTCGGTTTGGAGCTTGGAAATTCGTTTT	894
Db	579	rGly---ValIleMetLeuAlaCysValGly---GlyPheValAlaLysGlyLeuAsnLe	597
Qy	895	GGGAATCGATTTTAAAGAGGGTATGCTTTTACCTTTAAATCCAAAGAGCATGCATCAG	954
Db	597	uGlyIleAspPheGlnGlySerValIleGluValArgAlaLysGln---GlyGluAl	616
Qy	955	CGATGTTGCTCAATGCTGGCAAGTTGTGCATAACTACAGGAAGCTGCTCTTCTTC	1014
Db	616	aAspLeuAlaAspIleArg-----GluArgLeuAsnGlnLeuAsnLeuGly--	631
Qy	1015	TAGAGACTTCCGTATTCAACATTTTGGATTTCTCAGAAAAAGATCAAAATCTATTATTAGTGA	1074
Db	632	---GluIleGlnAlaGlnAsnPheGlySerProGlnAspValLeuIleArgIle----	648
Qy	1075	TAAAGCTTTAAGCTATATAAGCAGATACGAGCCTCTCTCTCTTAAATTTAAGCATCATGAG	1134
Db	649	-GlnAlaGlnAspGlyGlyGluAsnAlaGluGlnSerAlaIleThrLeu-----	664
Qy	1135	CTGGCGTATTGTGGGATTTGTGTACAGAAACAGCGCTAGATTCTCTACGGAACCTCTAA	1194
Db	665	-----ValArgGlyGlyLeu-----GluAspGly	672
Qy	1195	ACGAAACGCAAAATTTTGGTCAAAAGTAAGCAGCAAACTATCGAAGAAATATCGGTATCA	1254
Db	672	sTyArgPheArgArgValGluValValGlyProAlaValSerGlyAspLeuThrValTh	692
Qy	1255	GGCAGCATCGGCTTTTAGGAGCTTTGGCAATCATCTGCTCTATGATGATTTTGGCGTT	1314
Db	692	rSerThrIleGlyValLeuLeuAlaMetAlaIleMetValTrpValArgPh	712
Qy	1315	TGAATGCAATATGCTTTCAGTCCGCTATGCGCTTAAATTCATGACCTTTTGGCTACCTG	1374
Db	712	eGluTrpGlnPheAlaLeuGlyAlaValIleSerMetValHisAspValPheThr--	731
Qy	1375	TGCAGTCTTGTATTAGCACATTTCTTTTCAGAAAAATCAAAATAGATTTCGAAGCAT	1434
Db	732	-----IleGlyLeuPheValPheLeuGlyIleGluPheAsnLeuThrSeril	747
Qy	1435	TGGTGTCTTAATGACTGTTATGGGTTATTCATTAACAATCTTTGATCATTTTGTATCG	1494
Db	747	eAlaAlaIleLeuThrIleIleGlyTrpSerLeuAsnAspThrValIleTrpAspAr	767
Qy	1495	TATTCTGTAAGATCGCAAGCAACCTGTTTACCCCTATGATGCTTTTGTATGATGTC	1554
Db	767	gValArgGluAsnLeuArgTrpLysMetProLeuSerMetIleIleAspValSe	787
Qy	1555	CCCTCAAAAGCGTTGACCGCGCTCTCTCTTAAATTTTGCATTTTATATGACCATAGG	1614
Db	787	rLeuAsnGlnThrLeuSerArgThrIleLeuThrGlyLeuThrValLeuLeuAlaLeu	807
Qy	1615	AATGCTTTGTTTATAGCGCGCTCTCTCTTAAATTTTGCATTTTATATGACCATAGG	1674
Db	807	uAlaLeuThrLeuPheGlyGlyGluValIleArgSerPheThrPheAlaMetLeuPheGl	827
Qy	1675	GATTCTTCTAGGAATTTATGCTCTCTTTATATGACCACTCTCTGTTGTTG---TTTAT	1731



Db 827 yValGlyIleGlyValPheSerSerValTyrIleAlaIleProValIleuIleAlaPheIly 847  
 QY 1732 GGTCCGTAAGAAATCGC 1750  
 Db 847 sleuAArgProAspSerIlys 853

## RESULT 7

AB3337  
 protein translocase chain secf / protein translocase chain secf [imported] - Brucella me

C/Species: Brucella melitensis  
 C/Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002

C/Accession: AB3337

R.; DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova,  
 J.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A/Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A/Reference number: AD3252; PMID:11756688

A/Accession: AB3337

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-845 <KUR>

A/Cross-references: GB:AB008917; PIDN:AL51861.1; PID:G17982610; GSPDB:GN00190

A/Experimental source: strain 16M

C/Genetics:

A/Gene: BME10680

A/Map position: 1

Alignment Scores:

Pred. No.: 1,88e-54 Length: 845

Score: 715.50 Matches: 175

Percent Similarity: 52.87% Conservative: 129

Best Local Similarity: 30.43% Mismatches: 198

Query Match: 21.60% Gaps: 73

DB: 2 Indels: 14

US-09-868-987-1 (1-1864) x AB3337 (1-845)

QY 80 GCTGTAGTATTGATGCTTATATGTCACGACCCCTATTTTAAACGTCCTCCATTGAAA--- 136  
 Db 298 AlAlIleValIleuAspGlyValIleuSerAlaProValIleArgGluProIleThrGly 317  
 QY 137 AATCATGCGCATGCTCTCGGAAATTTACCGACGTGAAGGAGAACTCGCCGACAT 196  
 Db 318 GlySerGlyGlnIleSerGlySerPheThrValGlnAspThrValIleuSerIleu 337  
 QY 197 TTAATATCTGAGGAGATGCTTTGTTCCGAGTTCTCACTGAGAGACGATCTTCT 256  
 Db 338 LeuArgAlaGlyAlaIleuProAlaProIleuThrValIleGluGluArgThrValGlyPro 357  
 QY 257 GATCTGGGAAAAACAATGTACACAGGACATTTATCTCAGCATGCTGCTGGCAATG 316  
 Db 358 AspIleuGlyGlyAspAlaIleuSerMetGlyLeuMetThrGlyIleIleGlyPheLeu 377  
 QY 317 CTATATGTTTGAAGAGGTATATATGATTTTGGAGGCGTCATGCGTGGAGCGTGT 376  
 Db 378 ValAlaValPheIleLeuLeuLeuIleuIleuIleuIleuIleuIleuIleuIleu 397  
 QY 377 CTTCGATCTTTGCTTATCTGAGGAGCTCTACAGTATTGATGAGCCGACCTCTTG 436  
 Db 398 LeuLeuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 417  
 QY 437 TCAGAGCTCGTGGATGTTCTTCTGTAATGGAGTGGCCGATGATGCAATGTTCTTGA 496  
 Db 418 ProGlyIleAlaGlyIleIleIleuGlyIleGlyIleAlaValAlaAspAlaAsnIleLeu 437  
 QY 497 TTCGAAAGAAATCCGAGAGAAATTTATATGCTCAAGCTTTAAATAATCTGTAGAAAA 556  
 Db 438 AsnGluArgIleArgGluGluThrArgIleGlyLeuGlyAlaMetAlaIleuAspIys 457  
 QY 557 GGATATACCAAGCTTTGGAGCCATTTTGAATCTTAACTGATGACTGATTTGGCCTCA 616  
 Db 458 GlyPheIleSerAlaPheAlaThrIleValAlaAspAlaAsnValThrThrLeuThrAlaThr 477

QY 617 GCACTTCTTTCTCTAGATACAGGCGCTATTAAAGGTTGCTTGAACATGATTTA 676  
 Db 478 IleLeuLeuPheLeuPheGlyThrGlyProValArgGlyPheAlaValThrMetLeu 497  
 QY 677 GGAATTTCTCTCAATGTTTACGAGCTCTTTCATGACATTAATTTCTTCACTGTGG 736  
 Db 498 GlyIleAlaIleSerMetPheThrAspValThrLeuValArg---MetIleMetAlaTyr 516  
 QY 737 ATGATATAG-----ACC 748  
 Db 517 PheValArgArgArgIleuIleuValIleuIleuIleuIleuIleuIleuIleuIleu 536  
 QY 749 CAACATACAGATGATGATGATGAT---AGCTGGGAGATTAAGATGATTTCTTG 805  
 Db 537 GluHisThrAsnPheArgPheMetAsnAlaArgPheIleGlyIle----- 551  
 QY 806 AGAGATGCAAAAAACCTTGGGCTGTTCTGAGAGTGTCTTTTGAAGTTGCTGCT 865  
 Db 552 ---GlyValSerIleValIleuSerIleAla---SerIleIleuPheIleuProGly 569  
 QY 866 CTGGGCTTGGAGCTGGAATTCCTTTGGATGATTTTAAAGAGGATGCTT 925  
 Db 570 LeuAsnTyr-----GlyIleAspPheIleuGlyIleGlnAla 582  
 QY 926 ACCTTATCCAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 985  
 Db 583 GluIleSerThrSerGlnPro-----AlaAspLeuAlaGlnLeuArgAlaIleu 599  
 QY 986 CATTAACATACAGAGAGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1045  
 Db 600 -----GlyAlaLeuAsnLeuGlyGluValAlaLeuGlnThrAlaGlySer 614  
 QY 1046 TCAGAAAAAGTCAAAATCTATTT-----AGTATAAAGCTTTTAAAC 1087  
 Db 615 ProAsnGlnValIleuIleuArgValGlnArgGlnGlyGlyGluGluAlaGlnThrAla 634  
 QY 1088 TATATACAGATACAGAGCTCTCTCTTAAATTAAGATCAATGAGCTGCGCTATTGT 1147  
 Db 635 AlAlIleAsnIleuSerMetArgGluAlaValThrGluLeuAsp-----Pro 648  
 QY 1148 GGGATTTGTTTCAGAAACAGGCTAGATTTCTTACGAAACTTAAACGAAACGCAAA 1207  
 Db 649 GlyValAlaIle-----GluArgThrGluVal 657  
 QY 1208 TTTTGTCAAAAGTACAGCAAACTATGCAAGAAAATGCGTTATCAGGCGACATGG 1267  
 Db 658 ValGlyProIleValSerGlyGluLeuAlaArg-----SerGlyMetIleAla 673  
 QY 1268 CTTTATAGAGCTTTGGCAATCATCTTGTCTATGATGATGATGATGATGATGATGAT 1327  
 Db 674 ValIleuSerAlaValAlaMetLeuPheThrIleThrPheArgPheGluThrPhe 693  
 QY 1328 GCTTTCAGTCCGATGCGCTTAAATTCATGACCTTTTGGCTACCTGTGACGCTTGT 1387  
 Db 694 AlaLeuGlyAlaIleAlaThrLeuIleLeuAsp-----ThrThrIle 708  
 QY 1388 ATACACATTTCTTTGAAGAAATTAATCAATGATTTGCAAGCATGCTGCTTATATG 1447  
 Db 709 IleGlyPhePheAlaLeuThrGlnLeuAspPheAsnLeuThrAlaIleAlaIleLeu 728  
 QY 1448 ACTGATTTGGGATTTGATTAACAATCAATCTTATGATTTTGAATGATGATGATGAT 1507  
 Db 729 ThrIleIleGlyIleValIleAspIleValAlaValIleArgPheMetArgGluAsn 748  
 QY 1508 CGCAACGCAAGCTTTTACCCCTATGACATGTTTATGATGATGATGATGATGATGAT 1567  
 Db 749 MetArgLeuIleIleSerIleuIleuArgGluIleIleAspMetSerIleAsnGlnVal 768  
 QY 1568 TTCAGCCGACGATATGACAACTATCACTTATGATTTGTTGATGCTTTGTTGTT 1627  
 Db 769 LeuValArgCysIleIleThrSerMetThrPheLeuIleuIleuIleuIleuIleu 788  
 QY 1628 ATAGCGGCTCTCTGCTTATTTGATTTATATGACATGAGGATTTCTTCAAGA 1687



Db 789 TrpGlySerAlaValAlaHisAsnPheAlaValProMetLeuPheGlyValIleAla 808  
|||||.....  
Qy 1688 ACTTATCGTCTCTTATATTGACCACTCTGTGTGTTATG 1732  
|||.....  
Db 809 ThrSerSerIlePheIleAlaAlaProIleLeuLeuLeu 823  
|||.....  
RESULT 8  
D89946  
protein-export membrane protein SecDF [imported] - Staphylococcus aureus (strain N315)  
C;Species: Staphylococcus aureus  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C;Accession: D89946  
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura,  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A;Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A;Reference number: A89758; MUID:21311952; PMID:11418146  
A;Accession: D89946  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-759 <KUR>  
A;CROSS-references: GB:BA000018; PID:g13701435; PIDN:BAB42729.1; GSPDB:GN00149  
A;Experimental source: strain N315  
C;Genetics:  
A;Gene: secF

Alignment Scores:  
Pred. No.: 8.8e-52 Length: 759  
Score: 685.00 Matches: 171  
Percent Similarity: 53.01% Conservative: 137  
Best Local Similarity: 29.43% Mismatches: 197  
Query Match: 20.68% Indels: 76  
DB: 2 Gaps: 14

US-09-868-987-1 (1-1864) x D89946 (1-759)

Qy 104 CTCAGCAGCCCTATTAAACGTCCTCCATGAAATCATGCC---AGTGTCTCAGGGAAA 160  
|||.....  
Db 197 IleSerAlaAlaSerValAspGlnProIleAsnSerAspSerValGluIleSerGly 216  
|||.....  
Qy 161 TTTACCCACCGTGA-----GTGAGCAACTCGCTCAGATTTAAATCTGGA 208  
|||.....  
Db 217 PheLysGlyGlnGluGlyValLysAlaSerPheIleAlaGluLeuLeuAsnAlaGly 236  
|||.....  
Qy 209 CGCATGCTTTTGTTCGCGAGTCTCAGTCAAGAGAGATCTCTCTGATCTTGGGAAA 268  
|||.....  
Db 237 SerLeuProValAspLeuLysGluIleTyrSerAsnSerValGlyAlaGlnPheGlyGln 256  
|||.....  
Qy 269 AAACAATGTACACAGGCATTATCTCAGCATGCTGTGCGCTTGGCAATGCTTATTGTTTG 328  
|||.....  
Db 257 AspAlaLeuAspLysThrValPheAlaSerPheIleGlyValAlaLeuIleTyrLeuPhe 276  
|||.....  
Qy 329 ATGACGTATATTATAGATTGGAGGCTCATCGTTCGGAGCTGTCTCTGAATCTT 388  
|||.....  
Db 277 MetLeuGlyPheTyrArgLeuProGlyLeuValAlaIleAlaLeuThrThrTyrIle 296  
|||.....  
Qy 389 TTGCTATCTGGCGAGCTCTACAGTATTGGATGGCCACCTCACCTCTCAGGACTCGCT 448  
|||.....  
Db 297 TyrLeuThrLeuValAlaPheAsnPheIleSerGlyValLeuThrLeuProGlyLeuAla 316  
|||.....  
Qy 449 GGGATTGTCTGTGATGGGATGGCGGTAGATGCAATGTCTTGTATTTCGAAGAATC 508  
|||.....  
Db 317 AlaLeuValLeuGlyValGlyMetAlaValAspAlaAsnIleIleMetTyrGluArgile 336  
|||.....  
Qy 509 CGAGAGGAATTTTATCTCTCAAGTCTTAAATCTGTAGAAAAGGATATACCAAG 568  
|||.....  
Db 337 LysAspGluLeuArgileGlyArgThrIleLysGlnAlaPheSerLysAlaAsnLysSer 356  
|||.....  
Qy 569 GCTTTTGGACCATTTTGTATTCTAACTTACTACAGTATTGGCTCAGACTCTTTTC 628  
|||.....  
Db 357 SerPheLeuThrIlePheAspSerAsnLeuThrThrValIleAlaAlaValLeuPhe 376  
|||.....

Qy 629 TTCTAGATACAGGCGCTATTAAAGGCTTTGCTTTGACATTGATTTTAGGAATTTCTCT 688  
|||.....  
Db 377 PhePheGlyGluSerSerValLysGlyPheAlaThrMetLeuLeuLeuGlyIleLeuMet 396  
|||.....  
Qy 689 TCAATGTTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTG----- 733  
|||.....  
Db 397 IlePheValThrAlaValPheLeuSerArgPheLeuLeuSerLeuValSerSerAsn 416  
|||.....  
Qy 734 -----TGGATGAAATAGACCAACATACACATGTCATATGATGAAT 775  
|||.....  
Db 417 IlePheLysAsnGlnPheTrpLeuPheGlyValLysLysAsnLysArgHisAspIleAsn 436  
|||.....  
Qy 776 AGTTCCTGGGATA-----AAGCATGATTTCTTCGAGAGA 811  
|||.....  
Db 437 GluGlyValAspValHisAspLeuLysThrSerPheGluLysTrpAsnPheValLysLeu 456  
|||.....  
Qy 812 TGCAAAAACATTTGGGCTGTTTCTGGAAGTGTCTTTTCTTTAGTTCGCTCTCGG 871  
|||.....  
Db 457 AlaLysProLeuIleGlyValSerIleLeuValValGlyLeuValIleLeuTyr 476  
|||.....  
Qy 872 TTTGGAGCTGGAATTCGTTTGGGAATGGAATTTAAAGGAGGATATGCTTTTACCTTT 931  
|||.....  
Db 477 IlePheLysLeuAsn-----LeuGlyIleAspPheSerSerGlyThrArgValAspPhe 494  
|||.....  
Qy 932 AATCCAAAGAGCATGCATCAGCGATGTTGCTCAAAATGCGTGGCAAAAGTTGTCATAAA 991  
|||.....  
Db 495 GlnSerLysGln-----AlaIleThrGlnGlnLysValGluGlnVal 508  
|||.....  
Qy 992 CTACAGGAAGCTGCTTCTTCTTAGAGACTTCGCTATTCAACATTTGGATCTTCAGAA 1051  
|||.....  
Db 509 ValLysAspSerGlyLeuLysAlaAsp-----GlnIleGlnIleAsnGlyLysAspAsn 526  
|||.....  
Qy 1052 AAGATC---AAAATCTATTATTAGTATAAAAGCTTTAAGCTATATAAGCAGATACGAGCC 1108  
|||.....  
Db 527 LysValAlaThrValGlnPheLysAsp----- 536  
|||.....  
Qy 1109 TCTCTCTAAATTAACGATCATGAGCTGGCGTTATTGTGGGATTGTGTGAGAAACAGG 1168  
|||.....  
Db 537 -----LeuThrArgAlaGln 541  
|||.....  
Qy 1169 CCTAGATTTCTACGGAAACTCTAAACGAAACGCAAAATTTGG----- 1213  
|||.....  
Db 542 AspAsnLysLeuSerAspAsnIleLys-----SerLysPheGlyAspThrProGlnIle 559  
|||.....  
Qy 1214 TCAAGGTAAAGCAGCAAACTATCGAAGAAATGCTTATCAGGCGACCATCGGCTTTTA 1273  
|||.....  
Db 560 AsnThrValSerProIleIleGlyGlnGluLeuAlaLysAsnAlaMetLeuAlaLeuIle 579  
|||.....  
Qy 1274 GGAGCTTTGGCAATCATCTTGTCTATGTGAGTTTGGCTTTGAATGCAATATGCTTTC 1333  
|||.....  
Db 580 TyrAlaSerIleGlyIleIleIleTyrValSerLeuArgPheGluTrpArgMetGlyLeu 599  
|||.....  
Qy 1334 AGTCCGTATGCGCTTTTAAATTCATGACCTTTTGTGCTACCTGTGCGAGTCTGTGTTATAGCA 1393  
|||.....  
Db 600 SerSerValLeuAlaLeuLeuHisAspValPhe-----IleIleValAla 614  
|||.....  
Qy 1394 CATTTCTTTTGAAGAAATTCAAATAGATTTCGACCCATTGTGCTTTAATGACATGTA 1453  
|||.....  
Db 615 IlePheSerLeuPheArgIleGluValAspLeuThrPheIleAlaAlaValLeuThrIle 634  
|||.....  
Qy 1454 TTGGGTATTTCATTAAACAATCTTTCATCATTTTTCATCGTATTTCGTGAAGATCGCAA 1513  
|||.....  
Db 635 ValGlyTyrSerIleAsnAspThrIleValThrPheAspArgValArgGluAsnLeuGln 654  
|||.....  
Qy 1514 ---CGCAACTGTTTACC-----CCTATGATGCTTTTATGATGATGCTCCCTCAA 1561  
|||.....  
Db 655 LysValLysValIleThrThrThrGluGlnIleAspAspIleValAsnArgSerIleArg 674  
|||.....  
Qy 1562 AAGACGTTACGCGCAGGTAATGACACACAGCTACACTCTATCAGTGTGTTGTTAATGCTT 1621  
|||.....  
Db 675 GlnThrMetThrArgSerIleAsnThrValLeuThrValIleValValValAlaIle 694  
|||.....



```
Db 439 -----AlaPhePheIleValThrPhePheSerIleThrArgLeuGluValAspVal 455
Qy 1427 CAAGCCATTGGTCTTTAATGACTGTATTGGGGTATTTCATTAAACAATACCTTTGATCAT 1486
Db 456 ThrPheIleAlaAlaIleLeuThrIleIleGlyTyrSerIleAenAspThrIleValThr 475
Qy 1487 TTTGATCGTATTCTGAA-----CATCGCCAGCGAACCTGTTTACCCCTATGCAT 1537
Db 476 PheAspArgValArgGluHisMetLysLysArgLysProLysThrPheAlaAspLeuAen 495
Qy 1538 GTTTTAGTAAATGATGCCCTTCAAAAGACGTTTCAGCGCAGCGTAATGACACAGCTACA 1597
Db 496 HisIleValAenLeuSerLeuGlnGlnThrPheThrArgSerIleAenThrValIleThr 515
Qy 1598 ACTCTATCAGTCTTTGTTAATGCTTTTGTATAGCGCGCTCTCTGCTTTAAATTTGCA 1657
Db 516 ValValIleValValThrLeuLeuIlePheGlyAlaSerSerIleThrAsnPheSer 535
Qy 1658 TTTATTATGACCATAGGATCTCTTAGGAACCTTTATCGTCTCTTTATATATGACCACT 1717
Db 536 IleAlaLeuLeuValGlyLeuLeuThrGlyValTyrSerSerLeuTyrIleAlaAlaGln 555
Qy 1718 CTGTTGTTG 1726
Db 556 IleTrpLeu 558

RESULT 10
F75350
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75350
R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: F75350
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-768 <WHI>
A:Cross-references: GB:AF002022; GB:AF000513; NID:g6459590; PIDN:AAF11371.1; PID:g645959
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1822
A:Map position: 1

Alignment Scores:
Pred. No.: 2,48e-50 Length: 768
Score: 668.50 Matches: 181
Percent Similarity: 51.22% Conservative: 113
Best Local Similarity: 31.53% Mismatches: 223
Query Match: 20.18% Indels: 57
DB: 2 Gaps: 13

US-09-868-987-1 (1-1864) x F75350 (1-768)

Qy 77 ATGCGCTAGTGATTGACGGTTATATGTCAGCAGCCCTATTTTAAAGTCCTCCATTGAAA 136
Db 230 MetAlaValValLeuAspAspGlnIleGlnSerValAlaThrIleAenGlnArgLeuPhe 249
Qy 137 AATCATGCGCAGTGTCTCAGGGAATTTTACCACCGTGAAGTGACCAACTCGCCTCAGAT 196
Db 250 ArgAspIleGlnIleSerGlyAsnPheThrProGluAlaSerGlnLeuVal 269
Qy 197 TTAATACTGAGCGATGCTCTTTGTTCCCGAGGTTCTCAGTGAAGACGATCTCTCT 256
Db 270 LeuLysSerGlyAlaLeuProIleLysIleValThrAlaAlaGluArgSerIleGlyPro 289
Qy 257 GATCTTGGGAAAAACAATGTACACAGCGATTTATCTCAGCATGCTGTGGCTTGGCAATG 316
Db 613 TyrValGlyPheArgPheAspPheIleMetGlyLeuGlySerIleIleAlaAlaIleHis 632
```

```
Db 290 SerLeuGlyAlaAspAlaIleArgSerGlyAlaIleAlaLeuValGlyIleGlyLeu 309
Qy 317 CTATTGCTTTTGTAGTACGCTATATTAT---AGATTTGGAGCGTCATCGCTTCCGGAGCT 373
Db 310 ValPheValMetLeuPheAlaTyrTyrGlyLeuTrpPheGlyLeuValGlyAlaLeuGly 329
Qy 374 GTTCTTCTGAATCTTTTGGCTTAICTGGCGAGCTCTACAGTATTGTTGGATGCGCCACTCACC 433
Db 330 LeuLeuPheSerSerIleIleIleLeuGlyIleLeuGlyGlyPheGlyAlaThrLeuThr 349
Qy 434 TTGTCACGAGCTCGCTGGGATTTGTTCTTATGGGATGCGCGTAGATGCAAAATGTTCTT 493
Db 350 LeuProGlyIleAlaGlyLeuValLeuThrIleGlyAlaAlaValAspGlyAsnValIle 369
Qy 494 GTATTGAAAGAAATCCGAGAGGAATTTTATTGTCTCAAAAGTCTTAAATAATCTGTAGAA 553
Db 370 SerPheGluArgIleLysGluLeuAlaArgGlyLysGlyIleLysAenAlaIleGly 389
Qy 554 AAAGGATATACCAAGGCTTTTGGAGCCATTTTGAATCTTCTAATCTTACCTACAGTATTGGCC 613
Db 390 AlaGlyTyrGluHisSerThrAlaAlaIleLeuAspValAenAlaSerHisLeuLeuSer 409
Qy 614 TCAGCACTCTCTTTCTCTAGATACAGGCGCTATTAAAGGTTTGTGTTGACATTGATT 673
Db 410 AlaLeuAlaLeuTyrAenTyrSerThrGlyAlaValLysGlyPheAlaValThrLeuIle 429
Qy 674 TTAGGAATTTTCTCTCAATGTTTACCGCTCTTTTCATGACTTAATTTTCTTCATGCTG 733
Db 430 IleGlyValIleAlaSerThrPheSerAsnLeuValPheAlaLys---TrpPheMetGln 448
Qy 734 TGGATGAATTAAGCCCAACATACACAGTTGCATATGATGAATAAGTTCGTGGGATAAAG 793
Db 449 TrpLeuAlaGlnArgArgProAsn-----MetSerAlaProGlnTrpIleLys 464
Qy 794 CAA-----GATTTCTTGAGAGGATGCAAAATCTTTGGGCTGTTCTCGGAAGTGT 844
Db 465 HisThrHisPheAspPheMetLysProAlaLysValIleThrThrLeuSerValLeuLeu 484
Qy 845 TTTCTTTTAGTTCGTGCTGCTCGGTTTGGAGCTTGGAAATCCGTTTGGGAAATGAT 904
Db 485 AlaLeuAlaGlyAlaAlaLeuValAlaThrArgGlyLeuAen-----TyrGlyValAsp 502
Qy 905 TTTAAAGAGGATGATGCTTTTACCTTTTAAATCCAAAGAGCATGCGCATCAGCGATGTTGCT 964
Db 503 PheAlaProGlyThrThrLeuThrAlaArgValAspArgGlnValThrThrGlu----- 520
Qy 965 CAAATGCGTGCAGGAAGTTGTGCATAAACTACAGGAAGCTGTCTTCT----- 1012
Db 521 GlnLeuArgAenSerValIle-----GlyAlaGlyValSerLysValThrGly 536
Qy 1013 -----TCTAGAGACTTCCTGATTTCAACATTTTGGATCTTCAGAAAAGATC 1057
Db 537 GlnSerAlaThrIleGlnArgAspThrThrProGlyGlnGlnGlyGlnAsnPheThrVal 556
Qy 1058 AAAATCTATTATAGTAAAGCTTTAAGCTATATAAGCAGATACGAGCCTCTCTCTCTA 1117
Db 557 LysValProGluLeuAenAspAla-----GluValLysGlnIleGlyAlaAlaIleGly 574
Qy 1118 AAATTAACGATCATGACCTGCGGTTATTGTGGGATTTGTTGTCAGAAAACAGCCCTAGATT 1177
Db 575 LysLeuPro----- 577
Qy 1178 CTCTACGGAACCTCTAAACGAAACGCAAAATTTTGGTCAAAAGTAAGCAGCAAACTATCG 1237
Db 578 ---GlnGlyGlnValLeuAlaSerGluThrValGlyProAlaValGlyLysGluLeuThr 596
Qy 1238 AAGAAAATGCGTTTATCAGCGCACCATCGGGCTTTTAGGAGCTTTTGGCAATCATCTTGTCT 1297
Db 597 GlnLysThrIleTyrAlaValLeuLeuGly-----LeuGlyLeuLeuVal 612
Qy 1298 TATGTGATTTGCGCTTTGAATGGCAATATGCTTTTACGTCCGATGCGCTTAAATTCAT 1357
Db 613 TyrValGlyPheArgPheAspPheIleMetGlyLeuGlySerIleIleAlaAlaIleHis 632
```

```

QY 1358 GACCTTTGGCTACCTGTCAGCTCTGTTTATGACCACTTTCTTTGCAAAAAATTCA 1417
    |||:::
Db 633 AsyVal-----AlaIleAlaMetGlyLeuPheSerLeuGlyLeuGlu 647
QY 1418 ATGATTTGCAGCATTTGCTGTTTATGACTGTATGGGTATTCCTTAACAACT 1477
    :::::
Db 648 PheThrValAlaSerValAlaIleLeuThrLeuIleGlyTyrSerLeuAsnAspSer 667
QY 1478 TTGATCATTTTGTGATTCGTATTCGTGAAGATGCCAAGCAACTGTTTACCCCTATGCAT 1537
    :::::
Db 668 IleIleValSerAspArgIleArgGluAsnMetCysThrMetAlaGlyHisSerTyrArg 687
QY 1538 GTTTAGTAAATGATGCCCTTCAAAAGACGTTACAGCCGCAAGTATGACAAAGCTACA 1597
    :::::
Db 688 GluIleValAsnAlaAlaIleAsnGlnThrLeuSerArgThrValMetThrSerValSer 707
QY 1598 ACCTATACAGTTTGTGTAATGCTTTGTTTATGAGCGGCTCTGCTGTTTAAATTGCA 1657
    |||:::
Db 708 ThrMetLeuProLeuIleSerLeuLeuIlePheGlyArgProValLeuArgAspPheSer 727
QY 1658 TTTATTTGACCATAGGATTTCTTACAGCACTTTATGCTCTTTATATTGACACCACT 1717
    |||:::
Db 728 LeuIleLeuLeuValAlaGlyIleLeuValGlyThrTyrSerSerIleTyrIleValAlaPro 747
QY 1718 CTGTTGTTGTTTATGATC-----CGTAAAGAAAATCGCTCA 1753
    |||:::
Db 748 LeuValValTyrPheGluGluThrArgAspIysAsnArgAla 761

RESULT 11
AF3386
protein translocase chain secD / protein translocase chain secE [imported] - Brucella me
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AF3386
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
, Mazur, M.; Goleman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3386
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-775 <KOR>
A:Cross-references: GB:AE008917; PIDN:AAL52557.1; PID:gi7983044; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI1076
A:Map position: 1

Alignment Scores:
Pred. No. 5.15e-49 Length: 775
Score: 653.50 Matches: 166
Percent Similarity: 52.49% Conservative: 140
Best Local Similarity: 28.47% Mismatches: 204
Query Match: 19.73% Indels: 73
DB: Gaps: 15

US-09-868-987-1 (1-1864) x AF3386 (1-775)
QY 38 ACTGCTAATGACATATTCTGCAAAACCGTGATGCGTATGCTAGTATGACCGT 97
    |||:::
Db 219 ThrIleGlnGlyAsnGluAsnSer-----PheAlaIleValAlaAsn 233
QY 98 TATATGTCAGACGCCATTTTAAACGCCATG---AAAATCATGCCAGTGTCTCA 154
    :::::
Db 234 GlnValValSerAlaProThrValSerGlyProLeuAspThrSerGluLeuGlnIleIu 253
QY 155 GGAATAATTACCCACCGTGAAGTGAACAACCTCGCTCAGATTAAATCTGGACCGATG 214
    |||:::
Db 254 GlyAlaPheAspLeuGlnAlaAlaAsnAsnMetAlaValIleArgSerGlyAlaLeu 273
QY 215 TCTTTTGTCCGAGGTTTCTCAGTGAAGACGATCTCTTGATCTTGGGAAAAAACAA 274

```

```

Db 274 ProGlnAlaValThrValIleLeuGlnGluArgThrIleAlaSerAlaLeuGlyGluAspTyr 293
QY 275 TGACACAAAGCATTTATCTCAGCATCTGTGCTGGCAATGCTTATGTTGATGAC 334
    :::::
Db 294 AlaSerAlaValLeuAlaIleAlaLeuLeuAlaAlaLeuValAlaGlyLeuPheMetVal 313
QY 335 GTATATTATGATTTGAGGCCCTCATGCTTCCGGAGCTGTCTTGTGAATCTTTGCTT 394
    :::::
Db 314 LeuSerTyrGlyIleLeuGlyValIleAlaLeuValAlaLeuValAlaAsnIleIle 333
QY 395 ATGTCGACAGCTTACAGTATTTGATGATGCGSCACCTACACCTTACAGAGCTCGGAGATT 454
    |||:::
Db 334 LeuThrAlaValAlaLeuSerLeuIleGlyAlaSerIleSerLeuAlaSerIleAlaGlyLeu 353
QY 455 GTTCTTGCTATGCGGATGCGCTGATGACAAATGTTCTTATTCGAAGAATCTTCCTCAATG 514
    :::::
Db 354 ValLeuThrIleGlyLeuAlaValAlaAspAlaIleIleuIleTyrGluArgValAlaGly 373
QY 515 GAATTTTATGTCCTCAAAAGCTTTTAAATCTGTAGAAAAGATATACCAAGCTTTT 574
    :::::
Db 374 AspArgArgIysGlyTyrSerValValGlnAlaMetCysSerGlyPheTyrArgValAlaLeu 393
QY 575 GAGACATTTTGTGATCTGATGACTGACATGATGAGCTTACAGCACTCTTTCTTCTCA 634
    |||:::
Db 394 SerThrIleValAlaAspAlaAsnLeuThrThrLeuIleAlaIleValLeuIlePheLeu 413
QY 635 GATACAGGCTCTATTAAAGGTTTGTCTTGACATGTGATTTAGAAATTTTCTCTCAATG 694
    :::::
Db 414 GlySerGlyThrValHisGlyPheAlaLeuThrValAlaIleGlyIleGlyThrThrLeu 433
QY 695 TTTTACGCTCTTTTCATGATTAATTTTCTTCAATGCTGTCGATGAAT-----AAGACC 748
    |||:::
Db 434 PheThrThrLeuThrPheThrArgLeuLeuIleAlaGlnIlePvalArgThrAlaIysPro 453
QY 749 CAACATACACAGTTGATGATGATGAATTAAGTGTGGGATGAAGACT---GATTTGCT 805
    :::::
Db 454 LysGluValProLysArgAlaGlyLeuValAlaProThrValThrHisIleProPheMet 473
QY 806 AAGAGATGCAAAAACTTTGGGCTGTTCTGGAAGTGTCTTTTATGAGGTGC----- 859
    |||:::
Db 474 Arg-----LeuGlnPheValThrLeuGlyIleSerValLeuAlaCysAlaIle 489
QY 860 -----GTTGCTCTCGGTTTGGACCTCGGAATTCGTTTGGGAATGAT 904
    |||:::
Db 490 ValValAlaLeuPheValaIleGlyPheAsn-----TyrGlyIleAsp 504
QY 905 TTTTAAAGAGGGTATGCTTACCTTTAATCAAAA-----GAGCATGCGATCAGCGAT 958
    |||:::
Db 505 PheArgGlyGlySerMetValGluLeuGlnAlaArgAsnGlyAspAlaAsnLeuGluAsp 524
QY 959 GTTGTCTCAAAATGCGTGCAAAAGTTGTCATTAATCTACAGAGACGTGCTTCTTCTAGA 1018
    :::::
Db 525 IleAsnGlu-----ArgLeuAlaGluLeuAsnIleAspSer--- 536
QY 1019 GACTTCGCTATCAAACTTTGGATCTTCAGAAAAGATCAAAATCTATTATGATGAATA 1078
    :::::
Db 537 -----Ala 537
QY 1079 GCTTAACTATACATACAGATACAGACCTCTCTCTTAAATTAACATCATGACGTGG 1138
    :::::
Db 538 ArgValLeuProAlaLysSerProArgSerAlaLeuVal-----IleIleGlySer 554
QY 1139 CGTTATTTGTGGG-----ATTGTTCTCAAAAACAGCGCTAGATTTCTC 1180
    :::::
Db 555 GlnGluValAlaAspAlaProIleGlnThrValAlaValLysLeuArgGlyGluPheGlu 574
QY 1181 TACGGAACCTCT---AAGCAACGCAAAATTTTGTCGAAGAAGTATAGCAGCAACTATCG 1237
    :::::
Db 575 GlnAspTyrSerPheGlnArgValAspValValGlyProThrValSerGlnGlnLeuSer 594
QY 1238 AAGAAAATGGCTTATCAGGAGCATCGGCTTTTAGAGCTTTGGCAATCATCTTGCTC 1297
    :::::

```

Db 595 ArgAlaGlyValLeuAlaValIleLeuSerLeuLeuGly-----IlePheIle 610  
 Qy 1298 TATGTAGTTTCGCTTTGAATGCAATATGCTTTTCAGTGGCGTATGCGCTTTAAATCAT 1357  
 Db 611 TyrValTrpPheArgPheArgTrpGlnLeuAlaLeuGlyAlaValLeuSerThrLeuHis 630  
 Qy 1358 GACCTTTGGCTACTGTGCGAGTCTTGTGTTATAGACATTTCTTTTGAAGAAATCA 1417  
 Db 631 AspVal-----ValIleLeuSerGlyMetPheIleValPheArgMetGlu 645  
 Qy 1418 ATAGATTTCGAAGCATTTGGTCTTAAATGACTGTATTGGGTATTCATTAAACAATACT 1477  
 Db 646 PheAsnLeuTrpSerValAlaAlaIleLeuThrIleIleGlyTyrSerLeuAsnAspThr 665  
 Qy 1478 TTGATCATTTTGTATCGPATTTCGTAAGATCGCCAAAGCAACCTGTTTACCCCTATGCA 1537  
 Db 666 ValValIleTyrAspArgValArgGluAsnLeuArgArgTyrLysSerAlaProLeuPro 685  
 Qy 1538 GTTTTAGTAAATGATGCCCTTCAAAGAGCTTCAGCCGACGGTAATGACAACAGTACA 1597  
 Db 686 AlaIleIleAspAlaSerIleAsnGlnThrLeuSerArgThrLeuLeuThrSerPheVal 705  
 Qy 1598 ACTCTATCAGTTTGTAAATGCTTTTGTATAGCGGCTCTCTGCTCTTAAATTTGCA 1657  
 Db 706 ThrPheLeuAlaHisValProLeuTyrAlaPheGlyGlySerGluIleArgMetPheAla 725  
 Qy 1658 TTTATTATGACCATAGGATTTCTTCTAGGAATTTATCGTCTCTTTATATGCAACCACT 1717  
 Db 726 LeuAlaLeuSerValGlyIleIleValAlaSerTyrSerSerIlePheIleAlaAlaPro 745  
 Qy 1718 CTGTTGTTG 1726  
 Db 746 LeuLeuVal 748

RESULT 12  
 T34673  
 probable SecDF protein-export membrane protein - Streptomyces coelicolor  
 C;Species: Streptomyces coelicolor  
 C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
 R;Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.  
 submitted to the EMBL Data Library, December 1998  
 A;Reference number: Z21552  
 A;Accession: T34673  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-795 <SAU>  
 A;Cross-references: EMBL:AL034446; PIDN:CAA22394.1; GSPDB:GN00070; SCOEDB:SCIA9.24c  
 A;Experimental source: strain A3(2)  
 C;Genetics:  
 A;Gene: SCOEDB:SCIA9.24c

Alignment Scores:  
 Pred. No.: 5,99e-47 Length: 795  
 Score: 630.00 Matches: 181  
 Percent Similarity: 48.17% Conservative: 109  
 Best Local Similarity: 30.07% Mismatches: 212  
 Query Match: 19.02% Indels: 100  
 DB: 2 Gaps: 14

US-09-868-987-1 (1-1864) x T34673 (1-795)

Qy 74 CGTATGGCTAGTATGATGCGGTTATATGTCAGCAGCCCTATTTTAAACGTCCCATTTG 133  
 Db 220 ArgValAlaIleValLeuAspLysIleIleSerProGlnValAspProSerVal 239  
 Qy 134 AAA-----AATCATGCCAGTGTCTCAGGGAATTTTACCCACCGT 172  
 Db 240 SerCysGlyAlaGlyIleThrGlyGlySerThrGlnIleThrGlySerPheAspAspAla 259  
 Qy 173 GAATGACCAAACTCGCTCAGATTAAATCTGGAGGATGTCTTTTGTCCCGAGTT 232  
 Db 260 GluAlaArgGluLeuAlaLeuIleLysGlyAlaLeuProValProValGluThr 279

Qy 233 CTCAGTGAAGAGACGATCTCTCTCTGATCTTTGGGAAAAACAATGTACACAGGCATTATC 292  
 Db 280 IleGlnArgThrIleGlyAlaThrLeuGlyAspGluAlaIleAspAlaGlyAlaTrp 299  
 Qy 293 TCAGCATGCTGTGCTGGCAATGCTTATTGTTGATGAGCGTATATTATAGATTGGA 352  
 Db 300 AlaAlaValIleGlyThrAlaLeuThrAlaLeuPheIleIleValValTyrArgLeuMet 319  
 Qy 353 GGGCTCATCCCTCGGAGCTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 412  
 Db 320 GlyAlaLeuAlaThrValAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 339  
 Qy 413 TATTGGATCGCCACTCACCCTTGTCTAGGACTCGCTGGGATTTGTTCTTGTATGGGATG 472  
 Db 340 AlaValGlyAlaThrLeuThrLeuProGlyLeuAlaGlyPheValLeuAlaIleGlyMet 359  
 Qy 473 GCGTATGATCGAATGCTTGTATTTCGAAAGAAATCCGAGAGGAA-----TTTTATTG 526  
 Db 360 AlaValAspAlaAsnValLeuValPheGluArgAlaArgGluGluGlnAlaAlaArgThr 379  
 Qy 527 TCTCAAGTCTTAAATAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTT 586  
 Db 380 ArgProSerThrArgSerAlaLeuThrAlaGlyPheArgSerAlaPheSerAlaIleAla 399  
 Qy 587 GATTCTAATCTGACTACAGTATTGGCTCAGCACTTCTTTTCTCTAGATACAGGCCT 646  
 Db 400 AspSerAsnIleThrThrLeuIleAlaAlaLeuLeuPhePheLeuAlaSerGlyPro 419  
 Qy 647 ATTAAGGGTTTGTGTGACATGATTAGGATTTTCTCTCAATGTTTACGGCTCTT 706  
 Db 420 ValIleGlyPheGlyValThrLeuGlyIleGlyValLeuAlaSerMetValSerAlaLeu 439  
 Qy 707 TCTATGACTAAATTTTCTCTCTGCTGCTGATGAATAAGACCCCAACATACACATTCGAT 766  
 Db 440 ValIleThrArg-----ValLeuAlaGluPheAlaAlaSerArgProAlaValPhe 456  
 Qy 767 ATGATGAATTAAGTTCTGTTGGGATA----- 790  
 Db 457 ArgArgProArgIleThrGlyIleSerSerThrGlyProValArgAspAlaLeuLeuArg 476  
 Qy 791 AAGCATGATTCTTGTAGAGAGATGCAAAAACTTTGGCTGTTCTTCTGGAAGTGTTCCTT 850  
 Db 477 ArgAspProPheLeuMetArgArgProArgTrpLeuAlaAlaSerLeuIleValLeu 496  
 Qy 851 TTA-----GCTTGGCTGCTCTCGGGTTTGGAGCCTCGAATTCGTTTTCGGA 898  
 Db 497 ValValAlaGlySerGlyIleLeuValArgGlyLeuAsn-----PheGly 511  
 Qy 899 ATGGATTTTAAAGGAGGTATGCCCTTTTACCTTTTAAATCCA----- 937  
 Db 512 IleGluPheThrGlyGlyArgLeuIleGluTyrSerThrAlaThrGlnValAspProAsp 531  
 Qy 938 ---AAAGAGCATGCATCAGCATGTTGCTCAAAATGCTGGCAAGTGTGCTATAACTA 994  
 Db 532 ArgAlaArgAspAlaLeuAlaAspAlaGlyPheProArgAlaValVal----- 547  
 Qy 995 CAGGAAGCTGCTTCTTCTCTAGAGACTTCGCTATTCAACATTTGGATCTTCAGAAAAG 1054  
 Db 548 GlnSerSerGlyAspGlyAspLeuThrValArgThrGluGluLeuThrAspThrGluAla 567  
 Qy 1055 ATCAAAATATTATTAGTATGAAGCTTTTAAAGC-----TATACTAAGCAGATA 1102  
 Db 568 AlaThrVal-----ThrLysAlaValAlaGluLeuGlyGlyGluThrGluLysVal 584  
 Qy 1103 CGAGCCTCTCTCTTAAATAATTAACGATCATGCTGCGCTTATTGTTGGATTGTTGTCAGA 1162  
 Db 585 ArgAspGluLeu----- 588  
 Qy 1163 AACAGGCTAGATTCTCTACGGNAACTCTTAACGAACCGCAAAATTTTGGTCAAGGTA 1222  
 Db 589 -----ile 589

QY 1223 AGCAACAACTATGCAAGAAATGGTTATGACGGACCATCGGGCTTTTGAAGCTTGG 1282  
 Db 590 GlyProserLeuGlyGluGlnLeuArgAspAlaLeuIleAlaLeuGlyLeuAlaLeu 609  
 QY 1283 GCAATCATCTTGTCTATGATGAGTTGGCTTTGATGCAATATGCTTTCAGTCCGCTA 1342  
 Db 610 AlaIleGlnLeuAlaTylLeuAlaValArgPheArgLeuLeuPheGlyThrAlaIleVal 629  
 QY 1343 TGGCTTTAATTCATGACCTTTGGCTACCTGTGCACTGCTTTTATGACATTTCTTT 1402  
 Db 630 GlyAlaLeuAlaIleAspVal-----ValIleLeuValGlyValPheAla 644  
 QY 1403 TTGAAGAAATTCAAATGATTTTGCAAGCATGTGGCTTTAATGATGCTATTTGGGGTAT 1462  
 Db 645 TylPheGlyLysProIleAspGlyValPheLeuAlaIleLeuIleValIleGlyTyr 664  
 QY 1463 TCATTAACAAATCTTATGATGATTTTGTATGCTATTCGTGTA-----GATGTC 1510  
 Db 665 SerValAsnAspSerValValLeuPheAspArgIleArgGlyLeuGlyLysGlnArg 684  
 QY 1511 CAAGGAACTGTGTTACCCCTATGATGCTTTTATGATGATGCCCTTCAAAAGAGCTTC 1570  
 Db 685 LysAla-----ProPheAspArgLeuThrAsnAspAlaIleLeuGlnThrLeu 700  
 QY 1571 AGCCGACGCTATGACAAAGCAAGCTCACTCATGATGCTTTGTTATGCTTTGTTATA 1630  
 Db 701 ProArgThrValAsnThrGlyMetGlyAlaValIleIleLeuAlaSerLeuAlaIleLeu 720  
 QY 1631 GCGGCTCTCTGCTTAAATTTGATTTTATGACCAAGGATCTTCTTGAAGT 1690  
 Db 721 AlaAspAspSerLeuThrAspPheAlaLeuAlaLeuIleGlyValGlyValGlyThr 740  
 QY 1691 TTATGCTCTTTATATGACCAAGCTCTGTGTTGTTATGATGCTTGAAGAAATGCGC 1750  
 Db 741 TyrSerSerValPheThrAlaSerProLeuAlaIleGluLeuHisAsnArgAspThrGly 760  
 QY 1751 TCMAAA 1756  
 Db 761 SerArg 762  
 RESULT 13  
 A11627  
 protein-export membrane protein SecDF homolog l1n1562 (imported) - *Listeria innocua* (str  
 C;Species: *Listeria innocua*  
 C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C;Accession: A11627  
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.  
 D.; Jones, L.M.; Karet, U.  
 Science 294, 849-852, 2001  
 A;Authors: Kretz, J.; Kuhn, M.; Kunz, F.; Kurapkac, G.; Madueno, E.; Maitournam, A.; Ma  
 ok, C.; Schlueker, T.; Simoes, N.; Tiller, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
 A.; Title: Comparative genomics of *Listeria* species.  
 A;Reference number: AB1077; MUID:21537279; PMID:11679669  
 A;Accession: A11627  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-754 <GAP>  
 A;Cross-references: GB:AL592022; PIDN:CAC96793.1; PID:G16414049; GSPDB:GN00178  
 A;Experimental source: strain Clp11262  
 C;Genetics:  
 A;Gene: l1n1562

## Alignment Scores:

Pred. No.: 4.05e-46 Length: 754  
 Score: 620.50 Matches: 162  
 Percent Similarity: 47.13% Conservative: 117  
 Best Local Similarity: 27.36% Mismatches: 222  
 Query Match: 18.73% Indels: 91  
 DB: 2 Gaps: 13

US-09-868-987-1 (1-1864) x A11627 (1-754)

QY 98 TATATGTCAGACCCCTATTTTAAAGCTCCATTGAAATATCATGCCAGTGTCTCAGGG 157  
 Db 196 TyrLeuSerIleProAsnValIleSerSerValLeuAspThrAspLysValGluIleSerGly 215  
 QY 158 AAATTTACCAACCGGTGAAGTAGCAAACTGCGCTCAGATTTTAAATTTGAGGAGCTCT 217  
 Db 216 SerPheThrThrGlnGluAlaLysAspLeuAlaGluLeuLeuAsnSerGlyAlaLeuPro 235  
 QY 218 TTTGTTCCGAGGCTTCAGTAGAAGACATCTCTTCATCTTGGGAAAAACAATGT 277  
 Db 236 ValLysMetThrGluValTyrSerThrSerValGlyAlaGlnPheGlyAsnAlaLeu 255  
 QY 278 ACACAGGACATTAATCTCAGCATGCTGTGGCAATGCTTATGTTGATGAGCGTA 337  
 Db 256 GlnGluThrIleLeuAlaGlyIleIleGlyValMetAlaIlePheIlePheMetCala 275  
 QY 338 TATATGATTTGAGAGCGTCATGCTTCGGAGCTGTCTTGAATCTTTGCTTATC 397  
 Db 276 ValTyrArgLeuProGlyValIleAlaSerIleThrLeuValAlaTyrThrTyrLeuVal 295  
 QY 398 TGGGACCTCTACAGTATTTGATGCGCCACTGATCTTGCAGAGCTCGCTGGATGTT 457  
 Db 296 LeuLeuIleLeuSerLeuLeuAsnAlaThrLeuThrLeuProGlyIleAlaGlyLeuIle 315  
 QY 458 CTTCGTATGGGAGTGGCCGTGATGCAAAATGTTCTGTATTCGAAGATCCGAGAGAA 517  
 Db 316 LeuGlyIleGlyMetCalaValAspAlaAsnValIleThrTyrGluAlaGlyLysGluAlaPheArg 335  
 QY 518 TTTTATGCTCTCAAACTTAAATAATCTGTAGAAATAAGATATACCAAGCTTTTGA 577  
 Db 336 IleLysValGlyArgSerThrLysAlaIlePheGluValGlyGlyLysGluAlaPheArg 355  
 QY 578 GCCATTTTGAATCTTACATGATGATGATGATGATGATGATGATGATGATGATGAT 637  
 Db 356 AlaIleLeuAspGlyAsnLeuThrThrLeuIleValAlaValLeuPheTyrPheGly 375  
 QY 638 ACAGGCGCTTAAAGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 697  
 Db 376 ThrSerSerIleLysGlyPheAlaThrValLeuIleSerIleLeuValSerPheLeu 395  
 QY 698 ACGGCTCTTTCAGACTAAATTTTCTTCATGCTG-----TGATGAAT 742  
 Db 396 ThrAlaValTyrGlySerArgPheLeuLeuGlyLeuLeuValLysSerAsnThrLeuAsn 415  
 QY 743 AAG-----ACCAACATACACAGTGTGATGATGATGATGATGATGATGATGATG 784  
 Db 416 AsnLysProGlyPhePheAlaValLysArgLysAspIleHisAsnLeuHisGluGlyIle 435  
 QY 785 GCGATA-----AAGCATATTTCTTGAGAGATGCAAA 820  
 Db 436 AsnSerPheSerLeuLysThrHisPheAspArgTyrAspPheValHisHisArgLeu 455  
 QY 821 CTTTGGGCTGTTCTGAGAGCTGTTTCTTTAGTGTGCTGCTCGGATTTGAGGCC 880  
 Db 456 PheLeuSerIlePheAlaIleValIleValGlyIleValIleLeuSerIlePheLys 475  
 QY 881 TGAATTCCTTTGGGAATGATTTTAAAGAGGATGATGATGATGATGATGATGATGAT 940  
 Db 476 LeuAsn-----LeuGlyIleAspPheAlaSerGly-----ThrArg 487  
 QY 941 GAGCATGGCATCAGCGATGTGCTCAAAATGCGGCAAGATGTCATAAATCAGACGAA 1000  
 Db 488 AlaGluValThrAlaAsnGlnThrLeuThrGluThrGlnIleLysLysAspLeuAspAla 507  
 QY 1001 GCTGCTCTTTCTTGAAGACTTCGATTAACAATTTGATGATCTTCA----- 1048  
 Db 508 IleAspMetProSerAsnAspIleValPheGlnGlySerGlySerLysThrAlaValVal 527  
 QY 1049 -----GAAAGATCAAAATCTATTTAGTGAT 1075  
 Db 528 SerTyrLysGlyThrLeuSerGlnAsnAspValAlaLysPheLysAspTyrPheGluAsp 547  
 QY 1076 AAAGCTTAAGCTATCTAAGACAGATGCGAGCCCTCTCTCAAAATTAACGATCATGAGC 1135

```
Db 548 Lys-----TyrLysHisGluProSerIleSerThrValSerProThrVal----- 562
Qy 1136 TGGCGTTATTGTGGGATTGTTGTGAGAACAGCGCTAGATTCTCTACGAAACTCTAAA 1195
Db 563 -----GlyLysGluLeu 566
Qy 1196 CGAAACGCAAAATTTTGGTCMAAGGTAAAGCAGCAAACTATCGAAGAAATCGGTTATCAG 1255
Db 567 AlalysAenGlyPheTrpAla----- 573
Qy 1256 GCGACCATCGGGCTTTAGGA---GCTTTGGCAATCATCTGTCTCTATGTGAGTTTGGC 1312
Db 574 -----LeuGlyValAlaSerValLeuIleValLeuIleValLeuIleValAlaValArg 588
Qy 1313 TTTGAATGGCAATATGCTTTCAGTCGCGTATGCGCTTTAAATTCATGACCTTTTGGCTACC 1372
Db 589 PheGluPheTyrMetGlyIleAlaAlaIleLeuSerLeuLeuPheAsp----- 604
Qy 1373 TGTCAGTCTGTTTATAGCACATTTCTTTTGAAGAAATTCAAATAGATTTGCAAGCC 1432
Db 605 ---AlaPheIleIlePheIlePhePheSerValThrArgLeuGluValAlaSerLeuThrPhe 623
Qy 1433 ATTGGTGCTTTAAGTATGTTGGGTATTCATTAACAATACATCTTTGATCATTTTGTAT 1492
Db 624 IleAlaValLeuThrValIleGlyTyrSerIleAenAspThrIleValThrAlaAsp 643
Qy 1493 CGTATTGCTGAA---GATCGCGCAACGCAACCTGTTTACCCCTATGCTGTTTAA----- 1543
Db 644 ArgIleArgAenIleSerMetLysMetClnArgPheLysThrLysGluGluIleAlaAsp 663
Qy 1544 ---GTTAATGATGCGCTTCAAAGACGTTACGCGCAGCGGTAATGACAAACAGCTACAACT 1600
Db 664 AlaValAenLysAlaLeuArgGlnThrPheThrArgSerIleAenThrIleLeuThrVal 683
Qy 1601 CTATCAGTTTGTGTTAATGCTTTTGTATATAGGCGGCTCTCTGCTTTAATTTTGCATTT 1660
Db 684 IlePheThrValLeuAlaLeuValLeuPheGlySerGluSerIleLeuAenPheSerIle 703
Qy 1661 ATTAGTACCAGTAGGATCTTCTAGGAACCTTATGCTCTCTTTATATGACCACTCTG 1720
Db 704 AlaLeuLeuValGlyLeuValSerSerValPheSerSerIlePheMetAlaMetGlnLeu 723
Qy 1721 TTGTTGTTTATGTCGCGTAAGAAATCGTCAAAA 1756
Db 724 TrpTyrValPheLysAlaArgGlnLeuArgLysLys 735

RESULT 14
AG1265
protein-export membrane protein SecDF homolog lmo1527 [imported] - Listeria monocytogene
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AG1265
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Feihl, H.
D.; Jones, L.M.; Karst, U.
A;Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Varquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative Genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1265
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-754 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC99605.1; PID:g16410956; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo1527
```

```
Alignment Scores: 6.07e-46 Length: 754
Pred. No.: 618.50 Matches: 166
Score:
```

```
Percent Similarity: 46.79% Conservative: 111
Beat Local Similarity: 28.04% Mismatches: 224
Query Match: 18.67% Indels: 91
DB: 2 Gaps: 13

US-09-868-987-1 (1-1864) x AG1265 (1-754)

Qy 98 TATATGTCAGCAGCCCTATTTAAACGTCCCATTTGAAATATCATGCCAGTGTCTCAGGG 157
Db 196 TyrLeuSerAlaProAenValSerSerValLeuAspThrAspLysValGluIleSerGly 215
Qy 158 AATTTTACCACCGTCGAGTCAGCAAACTCGCCTCAGATTTAAATCTCGAGCGATGCT 217
Db 216 SerPheThrThrGluGluAlaLysAspLeuAlaGluLeuLeuAenSerGlyAlaLeuPro 235
Qy 218 TTTGTTCCCGAGTTCCTCAGTGAAGAGACGATCTCTTCATCTTGGGAAAAACAATGT 277
Db 236 ValLysMetLysGluValTyrSerThrSerValGlyAlaGlnPheGlyGlnAspAlaLeu 255
Qy 278 ACACAGGCATTTATCTCAGCATCTCTGGCTTGGCAATGCTTATTGTTTTCATGAGCGTA 337
Db 256 GlnGluThrIleLeuAlaGlyIleIleGlyValIleAlaIlePheIlePheMetMetAla 275
Qy 338 TATATAGATTTGAGCGCTCATCGCTTCGGAGCTGTTCTTCTGAATCTTTTCTTATC 397
Db 276 ValTyrArgLeuProGlyValIleAlaSerIleThrLeuValAlaTyrThrTyrLeuVal 295
Qy 398 TGGCAGCTCTACAGTATTGAGTGGCCACTCACCTTGTGAGGACTCGCTGGGATTTGT 457
Db 296 LeuLeuLeuLeuSerLeuLeuAenAlaThrLeuThrLeuProGlyIleAlaGlyLeuIle 315
Qy 458 CTTGCTATGCGGATGGCGTAGATGCAAAATGTTCTTGTATTTCGAAGAAATCCGAGAGAA 517
Db 316 LeuGlyIleGlyMetAlaValAlaAenValIleThrTyrGluArgIleLysGluGlu 335
Qy 518 TTTTATGTTCTCAAGTCTTAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGA 577
Db 336 IleLysValGlyArgSerThrLysAlaAlaPheGluValGlyGlyLysGluAlaPheArg 355
Qy 578 GCATTTTTCATTTTCACTTACCTGACATGATTTGGCTTCAGCACTTCTTTCTTCTTAT 637
Db 356 AlaIleLeuAspGlyAenLeuThrThrLeuValAlaAlaValLeuPheTyrPheGly 375
Qy 638 ACAGGCTATTAAAGGTTTGTCTTGACATGATTTTAGGAATTTTCTCTCAATGTTT 697
Db 376 ThrSerSerIleLysGlyPheAlaThrValLeuIleIleSerIleLeuValSerPheLeu 395
Qy 698 ACGGCTCTTTTCATGACTAAATTTTCTTCATGCTG-----TGATGAAT 742
Db 396 ThrAlaValTrpGlySerArgPheLeuLeuGlyLeuLeuValLysSerAenTrpLeuAen 415
Qy 743 -----AAGACCCCAACATACACAGTTGTCATATGATG 772
Db 416 AsnLysProGlyPhePheAlaValLysArgLysAspIleHisAenLeuHisGluGlyIle 435
Qy 773 AATAAGTTCGTGGGATAAAGCAT-----GATTTCTTGAGAGGATGCAAAAAA 820
Db 436 AsnSerPheSerLeuLysThrHisPheAspArgPheAspValLysHisHisArgLeu 455
Qy 821 CTTTGGGCTGTTTCTGGAAGTGTGTTTCTTTTAGTTTGGCTTCTCTCGGGTTTGGAGCC 880
Db 456 PheLeuSerIlePheAlaIleValIleValIleValIleLeuSerIlePheArg 475
Qy 881 TGGAAATCCGTTTGGGAATGGAATTTAAAGGAGGATGATGCCTTTTACCTTTTATCCAAA 940
Db 476 LeuAen-----LeuGlyIleAspPheAlaSerGly-----ThrArg 487
Qy 941 GAGCATGGCATCAGCGATGTTGCTCAAAATGCGGGCAAGCTTGTGCATAAATCTACAGAA 1000
Db 488 AlaGluValThrAlaAenGlnThrLeuThrGluThrGlnIleLysLysAspLeuAspThr 507
Qy 1001 GCTGGTCTTTCTTAGAGACTTCCCGTATTTCAAACATTTTGGATCTTCA----- 1048
Db 1001 GCTGGTCTTTCTTAGAGACTTCCCGTATTTCAAACATTTTGGATCTTCA----- 1048
```



Db 508 ILeaspMetProserAspIleValPheGlnGlySerGlySerThrAlaVal 527  
 Qy 1049 -----GAAAGATCAAAATCTATTTAGTAT 1075  
 Db 528 SerTyrIysGlyThrLeuSerGlnAspValAlaIysPheIysAsnTyrPheGluAsp 547  
 Qy 1076 AAACCTTTAGCTPACTCAACAGATACGACCTCTCTCCCAAAATTAATGACATGAGC 1135  
 Db 548 Lys-----TyrLysHieGlnProSerIleSerThrValSerProthVal----- 562  
 Qy 1136 TGGCGTTATGTGGGATTTGTGCAGAAACAGCCCTAGATTCTCTACGAAACTTAA 1195  
 Db 563 -----GlyLysGlnLeu 566  
 Qy 1196 CGAAACGCAAAATTTGGTCAAGGTAAGCAGCAAACTATGCAAGAAATCCGTTATAG 1255  
 Db 567 AlalysAsnGlyPheThrAla----- 573  
 Qy 1256 GCGACCATCGGGCTTTTGA-----GCTTGGCAATCATCTGCTATGAGTTTGGC 1312  
 Db 574 -----LeuGlyValAlaSerValLeuIleValLeuTyrIleAlaValArg 588  
 Qy 1313 TTGGAATGCAATATAGCTTTAGTCCGTATGCGCTTATATCATGACCTTTGGCTAAC 1372  
 Db 589 PheGluPheTyrMetGlyIleAlaIleLeuSerLeuLeuPheAsp----- 604  
 Qy 1373 TGTGAGTCTGTTTATAGCACTTTCTTTTGAAGAAATTCAAATGATTTGCAACC 1432  
 Db 605 --AlaPheIleIlePheIlePhePheSerValThrArgLeuGluValAspLeuThrPhe 623  
 Qy 1433 ATGTGCTTTAATGACGTATGGGATTCATTAACATTAATCTTATCATCTTTTAT 1492  
 Db 624 IleAlaIleValLeuThrValIleGlyTyrSerIleAsnAspThrIleValThrAlaAsp 643  
 Qy 1493 CGTATTCGTGAA--GATCGCCACAGCAACCTGTACCCCTATGCACTGTTTA----- 1543  
 Db 644 ArgIleAspAspIleSerMetLysMetGlnArgPheLysThrLysGluGlnIleAlaAsp 663  
 Qy 1544 --GTTAATGATGCCCTTCAAAAGACGTTACAGCCGACGGTATACACACAGCTCAACT 1600  
 Db 664 AlaValAsnLysAlaLeuArgGlnThrPheThrArgSerIleAsnThrIleLeuThrVal 683  
 Qy 1601 CTATCAGTTTGTCTTAACTTTTATAGCGCGCTCCTGCTTAAATTTGCACTT 1660  
 Db 684 IlePheThrValLeuAlaLeuValLeuPheGlySerGluSerIleLeuAsnPheSerIle 703  
 Qy 1661 ATTATGACCATAGGATTTCTTCTAGAACTTATGCTCTCTTATATGACACCACTCTG 1720  
 Db 704 AlaLeuLeuValGlyLeuValSerSerValPheSerSerIlePheMetAlaMetGlnLeu 723  
 Qy 1721 TTGTTGTTTATGTCCTTAAAGAAATGCTCAAA 1756  
 Db 724 TrpTyrValPheLysAlaArgGlnLeuArgLysLys 735  
 RESULT 15  
 C90594  
 protein-export membrane protein secD [imported] - Mycoplasma pulmonis (strain UAB CTIP)  
 C:Species: Mycoplasma pulmonis  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
 C:Accession: C90594  
 R:Chambud, I.; Hellis, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;  
 Nucleic Acids Res. 29, 2145-2153, 2001  
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul  
 A:Reference number: A99512; MUID:21267165; PMID:11353084  
 A:Accession: C90594  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-873 <KUR>  
 A:Cross-references: GB:AL445566; PID:gl4090074; PIDN:CA013832.1; GSPDB:GN00153  
 A:Experimental source: strain UAB CTIP  
 C:Genetics:  
 A:Gene: MYPU 6590  
 A:Genetic code: SGC3

Alignment Scores:  
 Pred. No.: 1,68e-37 Length: 873  
 Score: 522.50 Matches: 143  
 Percent Similarity: 47.778 Conservative: 135  
 Best Local Similarity: 24.578 Mismatch: 195  
 Query Match: 15.784 Indels: 109  
 DB: 2 Gaps: 18  
 US-09-868-987-1 (1-1864) x C90594 (1-873)  
 Qy 149 GTCTCAGGAAATTTACCCAGCTGAAGTGAAGCAAACTCCGCTCAGATTTAAATCTGCA 208  
 Db 298 IleThrGlyAsnLeuThrAlaSerSerAlaLysGlnLeuAlaLeuAspValAlaPheGly 317  
 Qy 209 GCGATGCTTTGTTCCCGAGGTTCCAGGAAGAGACGATCTTCTGATCTTGGGAA 268  
 Db 318 SerGlyAspTyrThrLeuGlnLeuLeuSerSerArgPheValAspAlaThrLeuGlyThr 337  
 Qy 269 AAACATGTACACAGCATTTATCTCAGCATGCTGTGGCAATGCTTATTTGTTTG 328  
 Db 338 GlnSerPheAsnTyrAlaIleIleAlaGlyIle-----IleSerPheIleValIleAla 355  
 Qy 329 ATGACGTATATATATAGATTGGA-----GGCGTATCGCTTCGGAGCTGTTCTTG 382  
 Db 356 LeuIleMetIleTyrAsnTyrGlyValLeuGlyIleAlaSerSerLeuSerIleGlyLeu 375  
 Qy 383 AATCTTTGCTTATCTGGCAGCTTACAGTATTTGGATGGCCACCTCACCTGTACGA 442  
 Db 376 TyrIleLeuLeuSerIleThrValPheThrTyrPheAsnGlyLeuTyrSerProAlaThr 395  
 Qy 443 CTCGCTGGATTTGTTCTTATAGGAGTGGCGGTATGATCAAAATGTTCTGTATTCGA 502  
 Db 396 IleAlaIleValIleAlaIleGlyMetSerValAspAlaAsnIleIleThrPheGlu 415  
 Qy 503 AGAATCCGAGAGAAATTTATGTCTCAAGTCTTAAATACTGTGAAGAAAGATAT 562  
 Db 416 ArgIleLysGlnGlnLeuLysTyrGlyAspSerLeuLysLysAlaThrLysAsnAlaHis 435  
 Qy 563 ACCAAGCTTTGGAGCCATTTTATGATCTTAACCTGACCTACAGTATGGCTCAGCACT 622  
 Db 436 ArgPheSerLeuSerThrIleLeuAspSerAsnIleThrThrLeuIleValSerGlyIle 455  
 Qy 623 CTTTTCTTCATGATCAGGAGCTTATTAAGGGTTGCTTGAACATTTATGAGATT 682  
 Db 456 LeuPheTyrPheGlyThrLysLeuIleArgGlyPheSerIleSerLeuIleLeuSerIle 475  
 Qy 683 TTCCTTCATGTTTACGGCTTTTTCATGACTAAATT----- 721  
 Db 476 IlePheIleLeuPheValMetLeuIleIleThrArgMetAlaThrSerLeuLeuIleGly 495  
 Qy 722 -----TTCCTCATGCTGTGATGATTAAGCAACCAATACACAGTTGCATATATGAT 775  
 Db 496 ThrGlyPhePhe-----AspAsnArgLeuTyrLeu----- 505  
 Qy 776 AAGTTCGGGAGATTAAGCATGATTTCTTGAGAGATGCAAAAACTT----- 823  
 Db 506 -----IleGlyValGlyArgSerPhePheThrLysGlyLysIleIleThrPhePhe 523  
 Qy 824 -----TGGCGTGT--TCGGAAGTGTCTTCTT 850  
 Db 524 GluLysLeuAsnTyrPheLysLeuSerTyrPheIleLeuPheGlyValPheVal 543  
 Qy 851 TTAGGTCGCTGCTTCGCGTTTGAGCCCTGGAATTCGTTTGGCA----- 898  
 Db 544 LeuIleAlaIleIleIle--TyrSerIlePheIleGlyLeuAsnGlyAsnPheLeuSer 562  
 Qy 899 -----ATGATTTTAAAGAGAGGATGCTTATCTTAAATCCAAAA-- 940  
 Db 563 GlyPheGluArgSerLeuGlnPheGlnGlyLysThrAsnIleSerPheGluLysSerSer 582  
 Qy 941 -----GAGCATGCAATCAGGATGCTTCAAAATGCGTGGCAAAAGTTGTCATTAACCTA 994









Db 409 SeranValleuIlePhegluArgVallysgluGluLeuArgLeuGIyAsnthrValser 428

Oy 542 AATCTGTAAGAAAAAGGATATACCAAGGCTTTGGAGCCATTGATCTACTGACT 601

Db 429 LysAlaIleGluLeuGIyPheLysArgThrLeuSerAlaValIleTyrAspThrIValThr 448

Oy 602 ACAGATTGGGCTCCAGCACTTCTTTCTTCCTAGATACAGGGCCATTAAAGGTTTGGT 661

Db 449 LeuLeuValAlaSerValIleLeuPheGlnPheGlySerGIyProValIleGlyPheAla 468

Oy 662 TTGACATTGATTTAGGAATTTCTTTCATGTTTACGGCTCTTTTCATGACATAATT 721

Db 469 ThrThrLeuAlaLeuGIyThrIleAlaSerPheIleSerAsnValIleTyrAlaLysVal 488

Oy 722 TTCTTCATGCTG 733

Db 489 PheLeuAspLeu 492

RESULT 21

H71326

probable protein-export membrane protein (secD) - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 28-Jul-2000

C:Accession: H71326

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwinn, E.; Nelson, K.E.; Paulsen, O.; Smith, T.F.; White, O.; Adams, J.; Peterson, J.; Kalk, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uitterlind, T.; McDowell, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:9832770; PMID:965876

A:Accession: H71326

A:Stratus: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-583 <COL>

A:Cross-references: GB:AE001219; GB:AE000520; NID:93322693; PIDN:AC65398.1; PID:G332265

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0410

C:Superfamily: protein export membrane protein secD

Alignment Scores:

Pred. No.:	4,63e-29	Length:	583
Score:	426.00	Matches:	92
Percent Similarity:	64.25%	Conservative:	50
Best Local Similarity:	41.63%	Mismatches:	77
Query Match:	12.86%	Indels:	2
DB:	1	Gaps:	2

US-09-868-987-1 (1-1864) x H71326 (1-583)

Oy 68 GGATGGCGTATGAGCTGTAGTGAATGACGGTATATANGTCAGACGCCCTATTAAACGTC 127

Db 343 GYAAGARGLeuAlaIleValserSepGIyLysIleArgSerAlaProAlaIleArgGlu 362

Oy 128 CGATG---AAAAATCATCCAGTGTCTCAGGGAATTTACCCACCGTAATGAGCAA 184

Db 363 ProIleThrAlaGIySerGIySerIleSerGIy---PheSerAlaGluGluAlaGlnAsn 381

Oy 185 CTCGGCTCGATTTAAATATCTGAGCGATGCTCTTTGTTCCGAGGTTTCTCAAGAG 244

Db 382 LeuLysThrAlaLeuArgSerAlaIlePheLeuAsnValAlaLeuGluIleGluAsnGlnGln 401

Oy 245 ACGATCTCTTCGATCTTGGGAAAAACAATGTACACAAGGATTAATCTCAGCATGCTGT 304

Db 402 ValValGIyAlaSerMetGIyGluGluSerIleArgGlnGIyThrArgAlaLeuValTyr 421

Oy 305 GCGTTGGCAATGCTATTTGTTGATGACCGTATATTAAGTATTTGAGAGGCGTCAATGCT 364

Db 422 GlyLeuCyAlaValLeuLeuPheMetLeuValIleTyrGIyGlnGluAlaGIyValAsnAla 441

Oy 365 TCGGGAGCGTGTCTTCTGATCTTTTGGCTATCTGGGAGCGCTCAAGATTTGGATGG 424

Db 442 CysValAlaGlnLeuLeuAsnLeuTyrIleMetPheGIyValLeuSerAlaPheAsnLeu 461

```

QY 425 CCACGACCTTGTTCAGGATCGCTGGATGTGCTTCGTCATGCGATGGCGATGATGCA 484
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 442 ThrLeuThrIleuSerSerIleAlaGlyMetIleuThrIleGlyMetAlaValAspAla 481
QY 485 AATGTTCTTGATTCGAAAGAAATCCGAGAGAAATTTTATTTGTCTCAAGCTTTAAAAA 544
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 482 AsnValIleValIlePheGluArgIleArgGluGluLeuAlaLeuGlyIleSerArgGlyAla 501
QY 545 TCTGTGAAAAAGGATATATCCAGGCTTTTGAGCGCATTTTATTTCTAATCTTGACATCA 604
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 502 AlaValCysSerGlyPheGluArgAlaPheThrPalaIleMetCysSerAsnValThrThr 521
QY 605 GTATGTGGCTTCAGCACTTCTTTCTTCCTAGATACAGGGGCTTTAAAGGTTTGTCTTG 664
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 522 PheIleAlaIleAlaLeuPheIleuSerValIleuGlyThrGlyProIleIleGlyPheAlaTyr 541
QY 665 ACATTGATTTTGAAGAAATTTTCTCTTCATATGTTTACGGCTCTTTTCATGACTAAATTTTTC 724
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 542 SerLeuAlaIleGlyValValSerSerValPheThrAlaLeuPheValSerArgLeuMet 561
QY 725 TTC 727
      ||| |||
Db 562 Phe 562

RESULT 22
F83169
secretion protein SecD PA3821 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83169
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
  Adam, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Li
  :; Lory, S.; Olson, M.V.
  Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pat
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F83169
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-620 <SNO>
A:Cross-references: GB:AE004799; GB:AE004091; NID:g9949981; PIND:AA607208.1; GSPDB:GNO
A:Experimental source: strain PA01
C:Genetics:
C:Gene: secD; PA3821
C:Superfamily: protein export membrane protein secD

Alignment Scores:
Pred. No.: 7,02e-29 Length: 620
Score: 424.00 Matches: 103
Percent Similarity: 53.71% Conservative: 49
Best Local Similarity: 36.40% Mismatches: 83
Query Match: 12.80% Indels: 48
Gaps: 5

US-09-868-987-1 (1-1864) x F83169 (1-620)

QY 62 AACCGTGAATGCGATGCTGCTGATG----- 88
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 357 AsnValIleValIlePheGluArgIleArgGluGluLeuAlaLeuGlyIleSerArgGlyAla 376
QY 89 -----ATTGACGCT----- 97
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 377 LysGluMetValAspIleValGluIleValAlaValProAlaPheIleGluIleIle 396
QY 98 TATATGTCAGAGAGCCCTATTAAACGTCCTTGAAGAAATCATGCAAGTCTCAGG 157
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 GlnIleIleSerLeuAlaThrIleGlnSerProIleGlyAsnGlnPheArgIleThrGly 416
QY 158 AATTTACCCACCGTGAATGTAGCAAACTGCGCTCGATTAAATTCGAGCGATGCT 217
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 LeuAspGlyProGlyGluIleSerGluLeuAlaLeuLeuLeuIleuArgAlaGlyGlyLeuAla 436

```

QY 218 TTTGTTCCGAGGTTCTCAGTGAAGAGAGGATCTCTTCTGATCTTGGGAAAAACAATGT 277  
Db 437 AlaProMetTyrPheAlaGluGluArgThrIleGlyProSerLeuGlyAlaAspAsnIle 456  
QY 278 ACACAAGCATTATCTCAGCATGCTGTGGCTTGGCAAGCTTATTGTTTTCATGAGCGTA 337  
Db 457 AlalysGlyIleAspAlaSerLeuTrpGlyMetLeuPheValSerLeuPheIleIleVal 476  
QY 338 TATTATAGATTGGAGCGTCATCGCTTCGGAGAGCTGTCTTCTGAAATCTTTTCTTATC 397  
Db 477 IleTyrArgPhePheGlyValIleAlaThrValAlaLeuAlaPheAsnMetValMetLeu 496  
QY 398 TGGCAGCTCTACAGTATTGGATGGCGCCATCACCCTGTGAGGACTCGCTGGGATGTT 457  
Db 497 ValAlaLeuMetSerIleLeuGlyAlaThrLeuThrLeuProGlyIleAlaGlyIleVal 516  
QY 458 CTGTGATGGGGATGGCGCTAGATGCAATGTTCTGTATTGCAAGAAATCCGAGAGGAA 517  
Db 517 LeuThrMetGlyMetAlaValAspAlaAsnValIleIlePheSerArgIleArgGluGlu 536  
QY 518 TTTTATTGTTCTAAAGTCTTAAAGTCTGTAGAAAAGGATATACCAAGCTTTTGGGA 577  
Db 537 LeuAlaAsnGlyMetSerValGlnArgAlaIleHisGluGlyPheAsnArgAlaPheThr 556  
QY 578 GCCATTTTGTATCTAACTTCACTACAGTATGGCTCAGCAGCTCTTTCTTCTCTAGAT 637  
Db 557 AlaIleLeuAspAlaAsnLeuThrSerLeuLeuValGlyGlyIleLeuTyrAlaMetGly 576  
QY 638 ACAGGGCTATTAAAGGTTTGTGATGATTGTTTGAAGATTTTCTTCTCAATGTTT 697  
Db 577 ThrGlyProValLysGlyPheAlaValThrMetSerLeuGlyIleIleThrSerMetPhe 596  
QY 698 ACGGCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGTGAATAAGCCCAACATACA 757  
Db 597 ThrAlaIleMetValThrArg-----GlyArgAspPhe-----Lys 616  
QY 818 AAATTTGG 826  
Db 617 LysLeuTrp 619  
RESULT 23  
I64056  
secretion protein secD - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 28-Jul-2000  
C:Accession: I64056  
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: I64056  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-616 <ITGR>  
A:Cross-references: GB:U32710; GB:I42023; NID:gi573200; PIDN:AAC21908.1; PID:gi573205; T  
C:Genetics:  
A:Gene: secD  
C:Superfamily: protein export membrane protein secD  
F:Keywords: inner membrane; protein export; transmembrane protein  
F:10-30/Domain: transmembrane #status predicted <TM1>  
F:457-473/Domain: transmembrane #status predicted <TM2>  
F:478-498/Domain: transmembrane #status predicted <TM3>  
F:503-519/Domain: transmembrane #status predicted <TM4>  
F:565-581/Domain: transmembrane #status predicted <TM5>  
F:587-606/Domain: transmembrane #status predicted <TM6>

Alignment Scores: 1.61e-27 Length: 616  
Pred. No.: 408.50 Matches: 87  
Percent Similarity: 65.85% Conservative: 48  
Best Local Similarity: 42.44% Mismatches: 59  
Query Match: 12.33% Indels: 11  
DB: Gaps: 2  
US-09-868-987-1 (1-1864) x I64056 (1-616)  
QY 137 AATCATCCAGTGTCCTCAGGGAATTT-----ACC 166  
Db 395 AsnValAlaThrIleGlnGlyArgPheGlySerAsnPheGlnIleThrGlyValAspSer 414  
QY 167 CACCGTGAAGTGACAACTCGCTCAGATTTAAATCTCGAGCGATGCTCTTTTGTTCCTC 226  
Db 415 IleAlaGluAlaHisAsnLeuSerThrLeuLeuLysSerGlyAlaLeuIleAlaProIle 434  
QY 227 GAGTTCTCAGTGAAGAGAGATCTCTTCTGATCTTGGGAAAAACAATGTACACAGGC 286  
Db 435 GlnIleValGluGluArgThrIleGlyProSerLeuGlyAlaGlnAsnValGluGlnGly 454  
QY 287 ATTATCTCAGCATGCTGTGCTGGCAATGCTTATTGTTGATGAGCGTATATTATAGA 346  
Db 455 IleAsnAlaSerLeuTrpGlyLeuValAlaValIleAlaPheMetLeuPheTyrTyrLys 474  
QY 347 TTTGGAGCGCTCATCGCTCGGAGCTGTCTTCTGATCTTTTGTCTTATCTTCTGGCAGCT 406  
Db 475 MetPheGlyValIleAlaSerPheAlaLeuValIleAsnIleValLeuValGlyLeu 494  
QY 407 CTACAGTATTG---GATGCGCACCTCCTTGTGAGACTCGCTGGGATTTTCTTGTCT 463  
Db 495 MetSerIleLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeuThr 514  
QY 464 ATGGGGATGCGCTGATGCAATGTTCTTGTATTCGAAAGATCCGAGAGGATTTTA 523  
Db 515 LeuGlyMetSerValAspAlaAsnValLeuIlePheGluArgIleLysGluIleArg 534  
QY 524 TTGTCTCAAAAGTCTTAAAAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCCATT 583  
Db 535 AsnGlyArgSerIleGlnAlaIleAsnGluGlyTyrAsnGlyAlaPheThrSerIle 554  
QY 584 TTGTATTTAACTTGACTACAGTATTGGCTCAGCATCTTTTCTTCTTCTTCTAGATACAGGG 643  
Db 555 PheAspAlaAsnLeuThrThrIleLeuThrAlaIleIleLeuTyrAlaValGlyThrGly 574  
QY 644 CCTATTAAAGGTTTGTGACATTGATTTAGGATTTTCTCTTCAATGTTTACCGCT 703  
Db 575 ProlleGlnGlyPheAlaIleThrLeuSerLeuGlyValAlaIleSerMetPheThrAla 594  
QY 704 CTTTTCATGACTAAA 718  
Db 595 IleThrGlyThrArg 599  
RESULT 24  
H81312  
protein-export membrane protein Cj1093c [imported] - Campylobacter jejuni (strain NCTC 11  
C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C:Accession: H81312  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A:Reference number: AB1250; MUID:20150912; PMID:10688204  
A:Accession: H81312  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-526 <PAR>  
A:Cross-references: GB:AL139077; GB:AL111168; NID:g6968444; PIDN:CB873348.1; PID:g696852;  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: secD; Cj1093c







QY 725 TTCATGCTGTG-----ATGATAAGACCCACAT 754  
 Db 393 LeuMet---TTPAlaIleAspMetCylLeuLeuValaSerHis 407

## RESULT 28

B87496

protein-export membrane protein SecD [imported] - Caulobacter crescentus

C/Species: Caulobacter crescentus

C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C/Accession: B87496

R;Nierman, W.C.; Deboy, R.T.; Paulsen, I.T.; Nelson, K.E.; Bisen, J.; Heidelberg, J.B.; Leub, M.T.; Debey, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolton, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A/Title: Complete Genome Sequence of Caulobacter crescentus.  
 A/Reference number: A87249; PMID:11259647  
 A/Accession: B87496  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-532 <STO>  
 A/Cross-references: GB:A805673; NID:G13423458; PIDN:AAK23966.1; GSPDB:GN00148  
 C/Genetics:  
 A/Gene: CC1991

## Alignment Scores:

Pred. No.: 5.39e-26 Length: 532  
 Score: 391.00 Matches: 85  
 Percent Similarity: 60.34% Conservative: 58  
 Best Local Similarity: 35.86% Mismatches: 90  
 Query Match: 11.81% Indels: 4  
 DB: 2 Gaps: 3

US-09-868-987-1 (1-1864) x B87496 (1-532)

QY 62 AACCGTGGCGGCGTATGATGACGGTTATATGTCAGACCCCTATTTA 121  
 Db 295 AsnValGlyLysArgPheAlaIleValLeuAspGlyAlaIleSerAlaProThrIle 314  
 QY 122 AAC--GTCCATTGMAAAATCATGCCAGTGTCTCAGGAAATTTACCCACCGTGMACTG 178  
 Db 315 AsnGlyAlaIleLeuGlyGlySerGlyIleIleThrGlySerPheThrAlaGluSerAla 334  
 QY 179 AGCAAACTCGCCTCAGATTAAATCTGGAGCGATGCTTTTGTCCGAGGTTCTCGT 238  
 Db 335 SerAspLeuAlaLeuLeuLeuArgSerGlyAlaLeuProAlaProLeuValGluGln 354  
 QY 239 GAAGAGCAGATCTCTTCATCTTGGGAAAAAACAATGATACAGACGATTATCTCACA 298  
 Db 355 GlnSerThrValGlyAlaGluLeuGlyAlaAspAlaValaGlyAlaIleSerThr 374  
 QY 299 TGCTGTGCTTGGCAATCTTATTTGTTGATGACGATATAT--AGATTTGGA 352  
 Db 375 LeuValAlaPheIleThrIleValValPheMetIleLeuSerGlyLeuLeuPheGly 394  
 QY 353 GGGCTCATCGCTCGGAGCTGTTCTTGAATCTTTGCTTATCTGGGACGCTTACAG 412  
 Db 395 Gly---IleSerValIleAlaLeuIleIleAsnGlyMetLeuIleValAlaIleMetSer 413  
 QY 413 TATTTGATGCGCCACCTCAGGATCGCTGGGATTTGTTCTTGATGGGAGT 472  
 Db 414 LeuThrGlnAlaThrLeuThrLeuProGlyIleAlaGlyLeuIleLeuThrLeuVal 433  
 QY 473 GCCGTAGATGCAATGTTCTGTATTCGAAGAATCCGAGAGAAATTTTATGTCTCA 532  
 Db 434 AlaValaPheAlaAsnValLeuIleLeuGlyArgMetArgAspGluAlaArgAlaGlyLys 453  
 QY 533 AGCTTAAAAAATCTGAGAAAAGATATACCAAGCTTTTGGAGCCATTTTATCT 592  
 Db 454 SerProIleLeuAlaIleAspAlaGlyPheSerArgAlaMetThrThrIleIleAspAla 473  
 QY 593 AACTGATACAGTATTTGGCTCAGCAGCTTCTTTCTTCCTAGATACAGGGGCTTTAA 652  
 Db 115 AACTGATACAGTATTTGGCTCAGCAGCTTCTTTCTTCCTAGATACAGGGGCTTTAA 652

Db 474 AsnValThrThrLeuValAlaIleAlaGlyIleMetPheAlaPheGlyAlaGlyProValArg 493  
 QY 653 GGGTTTCTTGCATGATGATTTAGAAATTTCTTTCATATGTTTACGCTCTTTTCATG 712  
 Db 494 GlyPheAlaThrPheLeuSerIleGlyValPheThrSerValPheThrAlaValLeuVal 513

QY 713 ACTAAATTTTCTTCATGCTGATGATATACCAATCATACACATG 763  
 Db 514 SerGlnLeuLeuIleGlyTyrPhePheArgAlaAlaArgProLysLysLeu 530

## RESULT 29

G81312

protein-export membrane protein Cj1092c [imported] - Campylobacter jejuni (strain NCTC

C/Species: Campylobacter jejuni

C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C/Accession: G81312

R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilli, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barr, Nature 403, 665-668, 2000  
 A/Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals h

A/Reference number: A81250; PMID:20150912; PMID:10688204  
 A/Accession: G81312  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-323 <PAR>  
 A/Cross-references: GB:AL139077; GB:AL111168; NID:G6968444; PIDN:CAM73347.1; PID:G6968

A/Experimental source: serotype O2, strain NCTC 11168

A/Genetic:

A/Gene: secF; Cj1092c

A/Superfamily: Escherichia coli preprotein translocase chain secF

## Alignment Scores:

Pred. No.: 6e-26 Length: 323  
 Score: 390.00 Matches: 98  
 Percent Similarity: 53.09% Conservative: 74  
 Best Local Similarity: 30.25% Mismatches: 96  
 Query Match: 11.78% Indels: 56  
 DB: 2 Gaps: 9

US-09-868-987-1 (1-1864) x G81312 (1-323)

QY 794 CATGATTTCTGAGAGGATGCAAAAACCTTGGCTGTTCT----- 835  
 Db 10 TyrAspPheMetArg-----MetArgPheAlaAlaIleSerLeuSerPheIleLeuPhe 27  
 QY 836 ---GGAAGTGTCTTTCTTATGAGTTGCGTTCCTCGGTTTGGAGCTGGAATTCGTT 892  
 Db 28 PheGlySerIleLeuLeu-----TTPAspArgGly 38  
 QY 893 TTG-----GGAATGATTTTAAAGAGGATATGCTTACCTTTATCCAAAAGCAT 946  
 Db 39 LeuIleThrGlyIleAspPheSerGlyGlyThrLeuValGlnLeu-----LysTyrGln 56  
 QY 947 GGCATCAGCATGTGTCGCAATGCGGCAAGATTGTGCATTAACATCAGAAAGCTGT 1006  
 Db 57 AsnAlaIleProIleThrGlnIleArg-----GluLeuGluAsnGln 71  
 QY 1007 CTTTCTTCTAGAGCTTCGATTCATCAAACTTGGATCTTCAGAAAAGATCAAAATCTAT 1066  
 Db 72 GlyThrPheGlnAsnLeuSerValThrGluPheGlySerAsnGluGluValThrIleArg 91  
 QY 1067 TTT-----AGTATTAAGCTTTAAGCTATACATCAGACATCAGAGCTCTCTCTTA 1117  
 Db 92 PheLeuGlySerAsnAspAsnValSerAspIleGlyIleHisIleSerThrLeuLeu 111  
 QY 1118 AATTAACGATCATGAGCTGCGCTTATTGAGATTGTTCAGAAAACAGGCTAGATT 1177  
 Db 112 LysAspThr----- 114  
 QY 1178 CTTCACGAAATCTTAACGAAACGAAATTTTGTCAAGGTACAGCAACATATCG 1237  
 Db 115 -----GlyLysPheGluValaArgAlaAspValaValGlyProLysValaGly 130





A:Residues: 1-615 <NR>

A/Cross-references: GB:AL590842; PIDN:CAC92424.1; PID:g15981127; GSPDB:GN00175

C:Genetics:

A:Gene: secD

C:Superfamily: protein export membrane protein secD

Alignment Scores:

Pred. No.:	2.52e-25	Length:	615
Score:	383.50	Matches:	86
Percent Similarity:	62.83%	Conservative:	56
Best Local Similarity:	38.05%	Mismatches:	81
Query Match:	11.58%	Indels:	3
DB:	2	Gaps:	2

US-09-868-987-1 (1-1864) x AE0387 (1-615)

```

QY 44 AATGACAAATATTCGTAACCCGTGATGCGTATGCTGATGATGACGTTATATG 103
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 375 SerGlyIysIysAspAlaAsn-----GlyArgSerIleLeuValIysGlnGluVal 392
QY 104 GTCAGCAGCCCTATTTAAACGTCCTCCATTGAAAAATCATGCCAGTCTCTCAGGAAATT 163
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 393 IleAsnValAlaThrIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleAsp 412
QY 164 ACCCAGCCTGAAGTGAAGCAAACTGCTCAGATTAAATCTGAGGAGTCTTTGTT 223
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 413 AsnProAlaGluAlaArgGlnLeuSerIleuLeuArgAlaGlyAlaLeuIleAlaPro 432
QY 224 CCCGAGGTTCTCAGTGAAGAAGACGATCTCTTCTGATCTGGGAAAAAACAATGACAA 283
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 433 IleGlnIleValIleGlnIleValThrIleGlyProThrIleuGlySerGlnAsnIleAlaGln 452
QY 284 GGCATTATCTCAGCAGTCTGTGCTGCTGCGCAATGCTTATGTTTATGAGCGTATATTAT 343
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 453 GlyLeuGluIaLacYsLeuTrpGlyLeuAlaValSerIleuPheMetValIaLysTrp 472
QY 344 AGATTGGAGCGCTCATCGCTTGGAGCTGCTTCTGATCTTTTCTTATCTGATCTGAGGA 403
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 473 ArgLysPheGlyValIleAlaSerThrAlaLeuMetAlaAsnLeuValIleValGly 492
QY 404 GCCTACAGTATTTG---GATGCCCACTCAGCTGTCAGGACTCGCTGGATTTGTTCT 460
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 493 ValMetSerIleuLeuProGlyAlaThrLeuThrMetProGlyIleAlaGlyIleValLeu 512
QY 461 GCTATGGGATGCGGTAGATGCAATGTTCTGATTCGAAAGATCCAGAGAAATT 520
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 513 ThrLeuAlaValAlaValAlaAspAlaAsnValLeuIleAsnGluArgIleLysGluGly 532
QY 521 TTATGTCCTCAAGTCTTAAAAATCTGTAGAAAAAGATPACCAAGCTTTGGAGCC 580
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 533 ArgAsnGlyArgThrIleGlnGlnAlaIleHisGluGlyTrpLysGlyAlaPheSerSer 552
QY 581 ATTTTATTTACTGACTGACATGATGTCGCTGACACACTCTTTCTTCTTCATGATACA 640
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 553 IleValAspAlaAsnIleThrThrLeuIleThrAlaIleIleLeuIleValAlaGlyThr 572
QY 641 GGGCTATTAAAGGTTGCTTTCGACATGATTTTGAAGATTTTCTTCATGTTTACG 700
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 573 GlySerIleLysGlyPheAlaIleThrThrAlaIleGlyValIaThrSerMetPheThr 592
QY 701 GCTCTTTTCATGACTAA 718
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 593 AlaIleValGlyThrArg 598

RESULT 32
B82285
protein-export membrane protein SecD VC0743 [imported] - Vibrio cholerae (strain N16961
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82285
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.U.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F
I., R.R.; Mekalanos, J.D.; Venter, J.C.; Fraser, C.M.

```

Nature 406, 477-483, 2000

A>Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A/Reference number: AB2035; MWID:20406833; PMID:10952301

A/Accession: B82285

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-617 <HEI>

A/Cross-references: GB:AE004160; GB:AE003852; NID:g9655181; PIDN:AAF93908.1; GSPDB:GN0

A/Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0743

A/Map position: 1

C:Superfamily: protein export membrane protein secD

Alignment Scores:

Pred. No.:	1.15e-24	Length:	617
Score:	376.00	Matches:	91
Percent Similarity:	57.32%	Conservative:	46
Best Local Similarity:	38.08%	Mismatches:	82
Query Match:	11.35%	Indels:	20
DB:	2	Gaps:	3

US-09-868-987-1 (1-1864) x B82285 (1-617)

```

QY 62 AACCTGAGAGGCGTATGCTGATGATGACGTTATATGTCAC-----AGC 112
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 362 AsnIleGlyLysLeuMetAlaThrValPheAlaGluTrpLysAspSerGlyLysArgSer 381
QY 113 CCATTTTAAACCTCCATGAAAAATCATGCGTGTCTCAGGAAAAATTCACCCCGT 172
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 382 ProGluGlyLysValIleLeuThrLysHisGluGluValIleAsnGlnAlaThrIleGln 401
QY 173 -----GAATGACCAA 184
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 402 SerAlaLeuGlyArgAsnPheArgIleThrGlyIleAspSerProAlaGluAlaHisAsn 421
QY 185 CTGGCTCAGATTAAATCTGAGGAGTCTTTGTTCCGAGTCTTCATGAAAG 244
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 422 LeuAlaLeuLeuLeuArgAlaGlyAlaLeuIleAlaProIleSerIleValIleGluLys 441
QY 245 ACATCTCTTCTGATCTTGGGAAAAAACAATGACAAAGCATTTATCTCAGCATGCTGT 304
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 442 ThrIleGlyProSerMetGlyGlnGlnAsnIleAspMetCylIleGlnAlaCylIleTrp 461
QY 305 GCGTGGCAATGCTTATGTTTGTATGAGCGTATPATTATGATTTGAGCGTATGCT 364
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 462 GlyMetValAlaValMetLeuPheThrValLeuTrpLysArgLysPheGlyMetIleAla 481
QY 365 TCGGAGCTGCTTCTGATGATCTTTGCTTATCTGAGCGTATPATTATGATTTGAGCGTATGCT 421
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 482 AsnIleAlaLeuMetAlaAsnLeuValIleIleGlyValMetSerMetIleProGly 501
QY 422 GCCCCTCAGCTTTCAGGAGTCTGCTGAGTATGTTCTTCTTTCATGAGCGTATGAT 481
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 502 AlaThrMetThrLeuProGlyIleAlaGlyIleValLeuThrValGlyMetAlaValAsp 521
QY 482 GCAATGTTCTTGTATTCGAAAGATCCGAGAGAAATTTTATGTTCTCAAGCTTAA 541
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 522 AlaAsnValLeuIlePheGluArgIleArgGluGluLeuArgLysAsnProGln 541
QY 542 AAATCTGTAGAAAAAGATPACCAAGCTTTGGAGCCATTTTATGTTACTGACT 601
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 542 GlnAlaIleHisGlnGlyTrpAlaAsnAlaPheSerThrIleAlaAspAlaAsnIleThr 561
QY 602 ACAGTATGGCTCAGACATCTTTCTTCATGATACAGGCGCTATTAAAGGTTGCT 661
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 562 ThrLeuIleThrAlaIleLeuPheAlaValGlyThrCylAlaIleLysGlyPheAla 581
QY 662 TTGACATGATTTTGAAGATTTCTCTTCATGTTTACGCTTTTTCATGACTAA 718
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 582 ValThrLeuSerIleGlyIleLeuThrSerMetPheThrAlaIleValGlyThrArg 600

RESULT 33

```

E83169  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: E83169  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Berman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: E83169  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-306 <STO>  
A:Cross-references: GB:AE004799; GB:AE004091; NID:99949981; PIDN:AAG07207.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
C:Superfamily: Escherichia coli preprotein translocase chain secF

Alignment Scores:  
Pred. No.: 3,75e-24 Length: 306  
Score: 369.50 Matches: 91  
Percent Similarity: 51.69% Conservative: 77  
Best Local Similarity: 28.00% Mismatches: 102  
Query Match: 11.16% Indels: 55  
DB: 2 Gaps: 10

US-09-868-987-1 (1-1864) x E83169 (1-306)

QY 779 TCTGTGGGTAAGCATGATTTCTTGAGAGGATGCAAAAACCTTGGCTGTTCTTGGGA 838  
Db 10 PheMetGlyLeuArgAsn-----ValAlaPhe 18  
QY 839 AGTGTCTTCTTTAGTGTGCTCTCGGTTTGGAGCTCGAATTCGTT----- 892  
Db 19 AlaValThrLeuIle----LeuThrValIleAlaLeuGlySerIlePheThrLysGlyLe 37  
QY 893 ---TTGGGAATGATTTTAAAGAGGGGTATGCCCTTTTACCTTTTAAATCCAAAGAGCATGGC 949  
Db 38 AsnPheGlyLeuAspPheThrGlyGlyThrLeuIleGluLeuThrTyrrGluInPro---- 56  
QY 950 ATCAGCATGTTGCTCAATCGCTGGCAAGTTGTGCATAAATACAGGAGCTGCTT 1009  
Db 57 ---AlaAspLeuGlyLysValArgGlyGlnLeuVal-----GlyAla 69  
QY 1010 TCTCTAGAGACTTCGCTATTCACATTTGGATCTTCAGAAAGATCAAAATCTATTT 1069  
Db 70 GlyTyrrGluAspAlaValValGlnSerPheGlyAspAlaArgAspValLeuValArgMet 89  
QY 1070 AGTGATAAAGCTTTAAGCTATATCAAGCAGATACGAGCTCTCTCTTAAATTAACGATC 1129  
Db 90 ProSerGluAspProGluLeuGlyLysValAlaThrAlaLeuGlnGlnAlaAsp--- 108  
QY 1130 ATGAGCTGGCGTTATTTGGGATTTGTTGTCAAAACAGGCGCTAGATTTCTCTACGGAAC 1189  
Db 109 -----AlaGly 110  
QY 1190 TCTAAACGAACGCAAAATTTTGTGTCGAAGTAAGCAGCAAACTATCGAAGAAATCGCT 1249  
Db 111 AsnProAlaAsnLeuLysArgValGluTyrrValGlyProGlnValGlyGluGluLeuArg 130  
QY 1250 TATCAGGCGACCATCGGCTTTTAGGAGCTTTGGCAATTCATTTGCTTATGTTGAGTTTG 1309  
Db 131 AspGlnGlyGlyLeuGlyMetLeuAlaLeuGlyGlyIleLeuTyrrValGlyPhe 150  
QY 1310 CGCTTGAATGGCAATATGCTTTCAGTCCGCTTAATTCATGACCTTTTGGCT 1369  
Db 151 ArgPheGlnTrpLysPheAlaLeuGlyAlaIleLeuSerLeuValHisAspAlaIle 170  
QY 1370 ACCTGTGAGTCTTGTATTAGCATTCTTTTGGGAAATTCAAATAGATTTGCAA 1429  
Db 1370 ACCTGTGAGTCTTGTATTAGCATTCTTTTGGGAAATTCAAATAGATTTGCAA 1429

Db 171 ValMetGlyValLeu-----SerPhePhe-----GlnValThrPheAspLeuThr 185  
QY 1430 GCCATTGCTGCTTTAATGACTGATTCGGGTATTTCATTAACAATCTTGTGATCATTTT 1489  
Db 186 ValLeuAlaAlaValLeuAlaValValGlyTyrrSerLeuAsnAspThrIleValIlePhe 205  
QY 1490 GATCGTATTGCTGAAGAT-----CGCAAGCGAACCTGTTTACCCCTATGCAT 1537  
Db 206 AspArgValArgGluAsnPheArgValLeuArgLysAlaAspLeuValGluAsnLeuAsn 225  
QY 1538 GTTTTAGTTAATGATGCCCTTCAAAAGACGTTCCAGCGCAGCGTAATACACACAGCTACA 1597  
Db 226 Ile-----SerThrSerGlnThrLeuLeuArgThrIleAlaThrSerValSer 241  
QY 1598 ACTCTATCAGTTTGTAAATGCTTTTATAGCGCGCTCTCTGTCTTTAAATTTTGA 1657  
Db 242 ThrLeuLeuAlaIleAlaAlaLeuLeuPheGlyGlyAspAsnLeuPheGlyPheSer 261  
QY 1658 TTATTATGACCATGAGGATCTTCTAGGAACCTTATCGTCTCTTATATATGACCACT 1717  
Db 262 IleAlaLeuPheValGlyValMetAlaGlyThrTyrrSerSerIleTyrrIleAlaAsnVal 281  
QY 1718 CTGTTGTTGTTTATG 1732  
Db 282 ValLeuIleIlePhe 286

RESULT 34  
F82429  
protein-export membrane protein SecD VCA0693 [imported] - Vibrio cholerae (strain N16961)  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 14-Sep-2001  
C:Accession: F82429  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bae, S.; Qin, H.; Dragoi, I.; Sellers, P. I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833; PMID:10952301  
A:Accession: F82429  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-612 <HEI>  
A:Cross-references: GB:AE004398; GB:AE003853; NID:99658102; PIDN:AAF96592.1; GSPDB:GN001:  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VCA0693  
A:Map position: 2  
C:Superfamily: protein export membrane protein secD

Alignment Scores:  
Pred. No.: 5,79e-24 Length: 612  
Score: 368.00 Matches: 79  
Percent Similarity: 59.17% Conservative: 63  
Best Local Similarity: 32.92% Mismatches: 94  
Query Match: 11.11% Indels: 4  
DB: 2 Gaps: 3

US-09-868-987-1 (1-1864) x F82429 (1-612)

QY 38 ACTGCTTAATCGACAATATCTCTGCAACCGTGGATGGCTATGCTGTAGTATGACGGT 97  
Db 370 ThrValTyrrArgGluTyrrLysThrAsnAla-----ArgGlyGluThrValArgSerGlu 387  
QY 98 TATATGCTCAGCAGCCCTTATTTAAAGCTCCCATTTGAAATCATGCCAGTGTCTCAGG 157  
Db 388 LysValIleSerValAlaThrIleGlnSerGlnLeuGlySerGlnPheArgIleThrGly 407  
QY 158 AAATTTACCCACCGTGAAGTGAGCAACTCGCTCAGATTTAAATCTGGAGCGATGTCT 217  
Db 408 AlaGlySerMetGluGluAlaGlnGlnLeuAlaLeuLeuArgAlaGlySerLeuThr 427  
QY 218 TTTGTTCCCGAGTTCTCAGTGAAGACGATCTTCTGATCTTTGGGAAAAAACAATGT 277

Db 428 AlaProvalThrIleValIGluGluArgThrIleGlyAlaSerLeuGlyGluAsnIle 447  
 QY 278 ACACAGGCATTTATCTCAGCATGCTGTCGCGATGCTTATTTGTTGATGAGCTA 337  
 Db 448 AlaAsnGlyPheAlaIleValAlaLeuGlyMetAlaMetThrLeuThrPheMetAlaLeu 467  
 QY 338 TATATATGATTGGAGGTCATGCGTTCGGAGCGTTCCTTCGTAATCTTTGGTTATC 397  
 Db 468 TrpIleArgValGluGlyTrpValAlaAsnValAlaLeuLeuValAlaMetAlaCysIleu 487  
 QY 398 TGGCAGCTCTACAGTATTTG---GATCGGCACCTACCTTGTGAGACTCGCTGGATT 454  
 Db 488 LeuGlyLeuIleAlaLeuLeuProGlyAlaValLeuThrLeuProGlyIleAlaGlyLeu 507  
 QY 455 GTTCTTGTATGGAGGCGCGTTCAGTCCAAATGTTTGTATTTGAAAGATCCGAGAG 514  
 Db 508 ValLeuThrValGlyMetAlaValAspThrAsnValLeuIlePheGluArgIleArgAsp 527  
 QY 515 GAATTTTATTTGTCCTCAAGCTTTAAATCTGTAGAAAAGATATACCAAGCTTTT 574  
 Db 528 LysLeuSerGluGlyArgAsnPheAlaGlnAlaIleAspThrGlyPheAspSerAlaPhe 547  
 QY 575 GGACCCATTTTGTATCTTAACCTGACTACAGTATGCGCTCAGCAGCTTTCTTCTCTA 634  
 Db 548 SerThrIleLeuAspAlaAsnIleThrThrMetIleThrAlaValIleLeuThrSerIle 567  
 QY 635 GATACAGGCGCTTAAAGGCTTTGCTTGAACATTGATTGAGATTCTTCTCAATG 694  
 Db 568 GlyAsnGlyProIleGlnGlyPheAlaLeuThrLeuGlyLeuGlyLeuThrSerMet 587  
 QY 695 TTTACGCGCTTTTTCATGACTTAAATTTTCTTC---ATCGTGTAGATGAATAGCCCAA 751  
 Db 588 PheSerGlyValPheAlaSerArgAlaLeuIleAsnLeuValTrpGlyArgAspThrArg 607

## RESULT 35

D83804 protein-export membrane protein secDF [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence revision 01-Dec-2000 #text change 15-Jun-2001  
 C:Accession: D83804

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: D83804  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-437 <STO>  
 A:Cross-references: GB:AF001511; GB:BA000004; NID:gl0173727; PIDN:BA04955.1; GSPDB:GNOC  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: secDF

Alignment Scores:  
 Pred. No.: 9e-24 Length: 437  
 Score: 365.50 Matches: 80  
 Percent Similarity: 56.92% Conservative: 68  
 Best Local Similarity: 30.77% Mismatches: 99  
 Query Match: 11.04% Indels: 13  
 Gaps: 4

US-09-868-987-1 (1-1864) x D83804 (1-437)

QY 20 CAGAGGGGATCAGCGCATCTAATGACAAATATCTGCACAAACCGTAGAGCGTATG 79  
 Db 182 GluGluGlyValAspSerTrpAlaGluGluAlaMetLysAlaSerProLysTrp----- 199  
 QY 80 GCTGTAGTATGATGACGTTTATATGCTCAGCAGCCCTATTTA-----AAGCTCCCATG 133  
 Db 200 -----IleSerAlaAlaSerValAsnGlnProLeuPheThrArgAspValMetIle 216  
 QY 134 AAAATATCATGCGATCTCAGCGAAATTTAATCCACCGTAGAAGTAGAGAACTGGCGTCA 193

Db 217 GluAsnArg-----SerGlyAspPheThrValGluGlnThrArgPheIleAlaAsp 233  
 QY 194 GATTAAATCTGGAGGATGCTTTTGTCCGAGTTCTTCAGTGAAGACATCTCT 253  
 Db 234 IleLeuAsnIleGlySerLeuProAlaLysLeuAspValLeuSerIleAsnSerValSer 253  
 QY 254 TCTGATCTTGGAAAAAACAATGATACACAGGATATTCCTCAGCATCTGCTGGCA 313  
 Db 254 AlaSerLeuGlyGluGlnAlaMetGluArgThrValIleArgPheIleGlyValAla 273  
 QY 314 ATGCTTATGTTTGTATGACGATATATATATGATTGGAGCGCTATCGCTTGGAGCT 373  
 Db 274 LeuIlePheLeuTrpMetLeuLeuTrpArgPheMetGlyMetIleAlaThrIleThr 293  
 QY 374 GTTCTTGTGATCTTTGCTTATCTGCGCAGCTCTACAGTATTTGATGGCCACTACC 433  
 Db 294 LeuThrValIleTrpIleTrpValLeuLeuIlePheAsnTrpMetAlaValLeuThr 313  
 QY 434 TTGTACAGACTCGCTGGATTTGTTCTTATGGGAGTGGCCGATGAGCAATGTTCTT 493  
 Db 314 LeuProGlyIleAlaAlaLeuIleLeuGlyValGlyMetAlaValAspAlaAsnIleIle 333  
 QY 494 GTATTCGAAAAGATCCGAGAGAAATTTTATTTGCTCAAGCTTTAAATAATCTGTAGA 553  
 Db 334 ThrTrpGluArgIleLysGluGluIleArgSerClyLysSerIleMetSerSerPheLys 353  
 QY 554 AAAGATATACCAAGCTTTTGGAGCCATTTTGATTTCTAATCTGATACATTTGGCC 613  
 Db 354 AlaIleSerArgArgSerLeuSerThrIleLeuAspAlaAsnIleThrIleLeuAla 373  
 QY 614 TCAGCAGCTCTTTCTTCCTAGATACAGGCGCTTAAAGGCTTTGCTTGACATGATT 673  
 Db 374 AlaSerValLeuTrpIleTrpGlyThrSerSerValGlnGlyPheAlaValMetLeuIle 393  
 QY 674 TTACGAATTTTCTTCAATGTTTACGCGCTTTTTCATGACATAATTTTCTTCATGCTG 733  
 Db 394 ValSerIleLeuMetSerPheValThrAlaValPheGlySerArgLeuLeuGlyLeu 413  
 QY 734 TGATGATTAAGACCCCAACATACACAGTTCATATGATGATTAAGTTTCGTGGATTAAG 793  
 Db 414 TrpValAsnSer-----ArgThrLeuAsnLysLysTrpArgLeuPheGlyValLys 430

## RESULT 36

AH0552

protein-export membrane protein SecF [imported] - Salmonella enterica subsp. enterica  
 C:Species: Salmonella enterica subsp. enterica serovar Typh

A:Note: this species has also been called Salmonella Typh

C:Date: 09-Nov-2001 #sequence revision 09-Nov-2001 #text change 27-Nov-2001  
 C:Accession: AH0552

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Church  
 et al.; Connerton, P.; Cronin, A.; Davis, P.; Davies, K.M.; Dowd, L.; White, N.; Fatta  
 , S.; Moutle, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica ser  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AH0552

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-323 <PR>

A:Cross-references: GB:AL513382; PIDN:CAD08864.1; PID:gl501677; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY0446

C:Superfamily: Escherichia coli preprotein translocase chain secF

Alignment Scores:  
 Pred. No.: 9.4e-24 Length: 323  
 Score: 365.00 Matches: 103  
 Percent Similarity: 49.68% Conservative: 53  
 Best Local Similarity: 32.80% Mismatches: 122  
 Query Match: 11.02% Indels: 36  
 Gaps: 11





C;Accession: H85536  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimailanta, B.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: H85536  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-323 <STO>  
A;Cross-references: GB:AE0051174; NID:g12513259; PIDN:AAG54756.1; GSPDB:GN00145; UWGP:Z05  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: secF  
C;Superfamily: Escherichia coli preprotein translocase chain secF

Alignment Scores:

Pred. No.:	1.15e-23	Length:	323
Score:	364.00	Matches:	102
Percent Similarity:	49.52%	Conservative:	53
Best Local Similarity:	32.59%	Mismatches:	124
Query Match:	10.99%	Indels:	34
DB:	2	Gaps:	9

US-09-868-987-1 (1-1864) x H85536 (1-323)

QY	794	CATGATTTCTTGAGAGGATGCAAAAACCTTTGGCTGTCTTCTGGAAGTGTCTTTCTTTTA	853
Db	17	TyrAspPheMetArgTyrAspTyrTrpAlaPheGlyIleSerGlyLeuLeuIleAla	36
QY	854	GCTTGCCTGCTCTCGGTTGGAGCTCGAATCCGTTTGGCAATGGATTTAAAGGA	913
Db	37	AlaIleValIleMetGlyValArgGlyPheAsn-----TrpGlyLeuAspPheThrGly	54
QY	914	GGGTATGCTTTACCTTTAATCCAAAAGATCGCATCGCGATGTTCCTCAATAGCT	973
Db	55	GlyThrValIleGluIleThrLeuGluIlyProAlaGluIleAspVal-----MetArg	72
QY	974	GGCAAGTTGTGCATAAATCTACAGAAAGCTGGTCTTTCTCTAGAGACTTCGCTATTCAA	1033
Db	73	-----AspAlaLeuGlnIlyAlaGlyPhe-----GluGluProMetLeuGln	86
QY	1034	ACATTTCGATCTTCAGAAAGATCAAAATCTATTTAGTGTAAAGCTTTAAGCTATACT	1093
Db	87	AsnPheGlySerSerHisAspIleMetValArgMetProAlaGluGlyGluThrGly	106
QY	1094	AAGCAGATACAGCCTCT---CTCCTAAATTAACGATCATGCTGGCGTATTGTGGG	1150
Db	107	GlyGlnValLeuGlySerGlnValIleIlyValIleAsnGluSer-----	121
QY	1151	ATTGTTGTCAAAACAGCGCTAGATTTCTCTACGGAAACTCTAAACGCAAAATTT	1210
Db	122	-----ThrAsnGlnAsnAlaAlaValIlyArg	130
QY	1211	TGTCAAAGTTAACGCAAACTATCGAAGAAATCGCTTATACGGCACCATCGGCTT	1270
Db	131	IleGluPheValGlyProSerValGlyAlaAspLeuAlaGlnThrGlyAlaMetAlaLeu	150
QY	1271	TTAGAGCTTTGGCAATCATCTCTCTATGTAGTTGGCTTTGATGCAATATGCT	1330
Db	151	MetAlaLeuLeuSerIleLeuValIlyValGlyPheArgPheGluIlyTrpArgLeuAla	170
QY	1331	TTCAAGTCGCTATCGCTTTAAATTCATGACCTTTTGGCTACCTGTGCGAGTCTTTTATA	1390
Db	171	AlaGlyValValIleAlaLeuAlaHisAspValIleIleThrLeuGlyIleLeuSerLeu	190
QY	1391	GCACATTTCTTTTGAAGAAATTCATATAGATTGCAAGCCATGTGCTTTTAAATGACT	1450
Db	191	PheHis-----IleGluIleAspLeuThrIleValAlaSerLeuMetSer	205
QY	1451	GTATTGGGTATTCATTAAACAAATCTTTGATCATTTTGTATCGTATTCGTAGAGAT---	1507
Db	206	ValIleGlyIlySerLeuAsnAspSerIleValValSerAspArgIleArgGluAsnPhe	225

QY	1508	CGCAACGCAACCTGTTTACCCCTATGCATGTTTATTAGTATGATGCCCTTCAAAAGACG	1567
Db	226	ArgIlyIleArgArgGlyThrProTyrGluIlePhe---AsnValSerLeuThrGlnThr	244
QY	1568	TTCAGCGCGCAGCGTAATGACAAACAGCTACAACTCTATCAGTCTTTGTTAAAGCTTTT	1627
Db	245	LeuHisArgThrLeuIleThrSerGlyThrThrLeuMetValIleLeuMetLeuLeu	264
QY	1628	ATAGGCGGCTCCTCTGCTTTTAAATTTTGCATTTTATATGACCATAGGATTTCTTAGGA	1687
Db	265	PheGlyGlyProValLeuGluGlyPheSerLeuThrMetLeuIleGlyValSerIleGly	284
QY	1688	ACTTATCGTCTCTTATATGACCACTCTCTGTTGTTG	1726
Db	285	ThrAlaSerSerIleTyrValAlaSerAlaLeuAlaLeu	297

RESULT 40

C81926

probable protein-export membrane protein NMA0812 [imported] - Neisseria meningitidis (ser:  
C;Species: Neisseria meningitidis  
C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 14-Sep-2001  
C;Accession: C81926  
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell  
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000

A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A;Reference number: A81775; MUID:2022556; PMID:10761919  
A;Accession: C81926  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-618 <PAR>  
A;Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84094.1; PID:g7379533  
A;Experimental source: serogroup A, strain Z2491  
C;Genetics:  
A;Gene: secD; NMA0812  
C;Superfamily: protein export membrane protein secD

Alignment Scores:			
Pred. No.:	1.3e-23	Length:	618
Score:	364.00	Matches:	80
Percent Similarity:	55.95%	Conservative:	61
Best Local Similarity:	31.75%	Mismatches:	105
Query Match:	10.99%	Indels:	6
DB:	2	Gaps:	3

US-09-868-987-1 (1-1864) x C81926 (1-618)

QY	23	GAGGGATACAGCGCACTGCTAATGGACAATAT---TCTGCAACCGTGGATGGCGTATG	79
Db	348	AspSerAlaGlyGlySerIlePheGlyGluLeuThrAlaAlaAsnValGlyLysArgMet	367
QY	80	GCTGTAGTGATT-----GACGGTTATATGCTCAGCAGCCCTTATTTAAACGTC	127
Db	368	AlaMetValLeuIleAspGingIlyLysSerGluValValThrAlaProValIleArgThr	387
QY	128	CCATTGAAA---AATCATGCCAGTCTCAGGGAATTTTACCACCGCTAGTGAGCAAA	184
Db	388	AlaIleThrGlyArgValGluIleSerGlySerMetThrThrAlaGluAlaAsnAsp	407
QY	185	CTCGCTCAGATTTAAATCTGGAGCGATGCTTTTGTCTCCGAGGTTCCTCAGTGAAGAG	244
Db	408	ThrSerLeuLeuLeuArgAlaGlySerLeuAlaAlaProMetGlnIleValGluArg	427
QY	245	ACGATCTCTCTCATCTTGGGAAAAAACAATGTACAAAGCATTATCTCAGCATGCTGT	304
Db	428	ThrIleGlyProSerLeuGlyLysGluAsnIleGluIlyGlyPheHisSerThrLeuTrp	447
QY	305	GGCTGGCAATGCTTATTGTTTTCATCAGCGTATATTATATAGATTTCGAGCGCTCATCGCT	364
Db	448	GlyPheAlaIleValAlaAlaPheMetValValTyrTyrArgLeuMetGlyPhePheSer	467
QY	365	TCGGGAGCTGCTTCTCTGAATCTTTGCTTATCTGGGAGCTCTACAGTATTTGGATCGC	424







Db	12	AspPheMetArgTrpSerLysPheAlaPheAlaLeuSerLeuValMetIleAlaAlaSer	31
Qy	842	GTITTTTCTTTTAGTGTTCGCTCTCGGGTTTGAGAGCCTCGAAATCCGTTTTCGGGAATG	901
Db	32	IlePheThrLeuSerThrLysTrpLeuAsn-----Trp-----GlyLeu	44
Qy	902	GATTTTAAAGAGGGTATGCGCTTTACCTTTTAATCCAAAGAGCATGGCATCGGATGTT	961
Db	45	AspPheThrGlyThrLeuIle-----GluValGlyPheGluGlnPro	59
Qy	962	GCTCAAAATCGTGGCAAGTTGTGCATAAATACAGGAAGCTGCTCTTCTCTAGAGAC	1021
Db	60	AlaAsnLeuGlu---GlnIleArgSerAlaLeuGluAlaLysGlyPheGly-----Asp	76
Qy	1022	TTCCGTATTCAAAACATTTCGATCTTCAGAAAAAGATCAAAATCTAT-----	1066
Db	77	AlaThrValGlnAsnPheGlySerAlaArgAspValMetValArgLeuArgProArgAsp	96
Qy	1067	---TTTAGTGATAAAGCTTTAAAGCTATATAAGCAGATACGAGAGCTCTCTCTAAATA	1123
Db	97	AspMetAlaGlyGluAlaLeuGly-----AsnGlnIleLeuAlaAlaIleLysGluGly	114
Qy	1124	ACGATCATGAGCTGGCGTTATTGTGGATTGTGTGCAGAAACAGCCTAGATTCTCTAC	1183
Db	115	Thr-----GlyGlyAsnValGluMetArgIleGluPhe-----	126
Qy	1184	GGAAACTCTAAACGAAACGCAAAATTTTGGTCAAAGGTAAAGCAGCAACTATCGAAGAA	1243
Db	127	-----ValGlyProAsnValGlyAspGlu	134
Qy	1244	ATCGGTTATCAGCGACCATCGGGCTTTTAPAGAGCTTTTGGCAATCATCTTGCTCTATGTG	1303
Db	135	LeuThrGluAlaGlyGlyLeuAlaIleLeuValSerLeuLeuCysIleLeuLeuTyVal	154
Qy	1304	AGTTTGGCTTTCAAGTGGCAATATGCTTTCAGTGCCTGATCGGCTTTAATTCATGACCTT	1363
Db	155	SerValArgPheGluTrpArgLeuAlaAlaGlyAlaValLeuAlaLeuAlaHisAspVal	174
Qy	1364	TTGGCTACCTGTGCAGTCTGTTTATAGCACATTTCTTTTGAAGMAAATTCAAATAGAT	1423
Db	175	IleIleThrLeuGlyIle-----PheSerIleLeuGlnIleGluValAsp	189
Qy	1424	TTGCAAGCCATTGGTGCTTTTAATGACTGTATTGGGTATTTCATTAACAATACTTTGATC	1483
Db	190	LeuThrIleValAlaAlaLeuLeuThrValValGlyTyTrpSerLeuAsnAspThrIleVal	209
Qy	1484	ATTTTTCATCGTATTTCGTGAAGAT---CGCCAAAGCAACCTGTTTACCCTATGTCATGTT	1540
Db	210	ValPheAspArgIleArgGluAsnPheArgLysMetArgLysGluGluProAlaGluIle	229
Qy	1541	TTAGTTAATCATGCCCTTCAAAAGAGCTTCACGCCACGGTAATGACAAACAGCTACAAC	1600
Db	230	Met---AsnSerSerIleThrGlnThrLeuSerArgThrLeuIleThrSerGlyThrThr	248
Qy	1601	CTATCAGTTTTGTTAATGCTTTTGTTTATAGCGGCTCTCTCTGTTCTTTAATTTTGCATTT	1660
Db	249	LeuPheValValIleAlaLeuPheThrGlnGlyAlaMetIleHisGlyPheAlaLeu	268
Qy	1661	ATTATGACCATAGGATTTCTCTTAGGAATTTTATCGTCTCTTTATATGACCAACCTCTG	1720
Db	269	AlaLeuLeuLeuGlyIleThrValGlyThrTyTrpSerSerIleTyValAlaSerAlaLeu	288
Qy	1721	TTGTTG 1726	
Db	289	AlaLeu 290	
RESULT 43			
H64769			
preprotein translocase chain secD [validated] - Escherichia coli (strain K-12)			
N;Alternate names: protein-export membrane protein secD; secretion protein A			
C;Species: Escherichia coli			
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 01-Mar-2002			

C:Accession: H64769; JQ0696; S12301  
R:Blattner, F.R.; Plunkett III, G.;  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: H64769  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-615 <BLAT>  
A:Cross-references: GB:A8000147; GB:U00096; NID:G1786603; PIDN:AACT3511.1; PID:G1786609;  
A:Experimental source: strain K-12, substrain MGL655  
R:Gardel, C.; Johnson, K.; Jacq, A.; Beckwith, J.  
EMBO J. 9, 3209-3216, 1990  
A:Title: The *secD* locus of *E. coli* codes for two membrane proteins required for protein export  
A:Reference number: JQ0693; MUID:91006014; PMID:2170107  
A:Accession: JQ0696  
A:Molecule type: DNA  
A:Residues: 1-77, 'S', '79-154', 'A', '156-615' <GAR>  
A:Cross-references: GB:X56175; NID:G42929; PIDN:CAA3634.1; PID:G581230  
R:Gardel, C.; Johnson, K.; Jacq, A.; Beckwith, J.  
EMBO J. 9, 4205-4206, 1990  
A:Reference number: S12298; MUID:91065354; PMID:2249673  
A:Contents: erratum  
A:Accession: S12301  
A:Molecule type: DNA  
A:Residues: 1-77, 'S', '79-154', 'A', '156-615' <GA2>  
A:Cross-references: EMBL:X56175; NID:G42929; PID:G581230  
R:Pogliano, K.J.; Beckwith, J.  
J. Bacteriol. 176, 804-814, 1994  
A:Title: Genetic and molecular characterization of the *Escherichia coli secD* operon and  
A:Reference number: A36969; MUID:94131960; PMID:7507921  
A:Contents: annotation; membrane topology  
C:Genetics:  
A:Gene: *secD*  
A:Start codon: GTG  
C:Complex: heterohexamer; chains *secY* (PIR:Q08CSY), *secE* (PIR:VXECSE), *secG* (PIR:S40402)  
ded trimeric complex of *secY*, *secE* and *secG* and the peripheral *secA* protein; the protein  
associated to form *SecYEGFP* is a hexameric integral membrane domain of the pre-protein  
C:Function:  
A:Description: the *secD* protein is a transmembrane component of the protein export complex  
C:Superfamily: protein export membrane protein *secD*  
C:Keywords: inner membrane; protein export; transmembrane protein  
F:10-30/Domain: transmembrane #status predicted <TM1>  
F:31-455/Domain: periplasmic #status predicted <PP1>  
F:456-472/Domain: transmembrane #status predicted <TM2>  
F:477-497/Domain: transmembrane #status predicted <TM3>  
F:498-501/Domain: periplasmic #status predicted <PP2>  
F:502-518/Domain: transmembrane #status predicted <TM4>  
F:564-580/Domain: transmembrane #status predicted <TM5>  
F:581-585/Domain: periplasmic #status predicted <PP3>  
F:586-605/Domain: transmembrane #status predicted <TM6>

Alignment Scores:  
Pred. No.: 3.23e-23 Length: 615  
Score: 359.50 Matches: 80  
Percent Similarity: 62.3% Conservative: 61  
Best Local Similarity: 35.4% Mismatches: 82  
Query Match: 10.8% Indels: 3  
DB: 1 Gaps: 2

US-09-868-987-1 (1-1864) x H64769 (1-615)

Qy 44 ATGGACAATATCTGCAACCGTGGATGGCTGTAGTGATTGACGGTTATATG 103  
Db 375 SerGlyLyLeAspAlaAsn-----GlyArgAlaValLeuValysGlnGluVal 392  
Qy 104 GTCAGCAGCCCTATTTTAAACGTCCTCCATGAAATCATGCCAGTGTCTCAGGGAATTT 163  
Db 393 IleAsnIleAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleAsn 412  
Qy 164 ACCCAACCGTGAAGTAGCAGCAACTCGCTCAGATTAAAAATCTGAGCGCATGCTTTTGT 223

```

Db 413 AsnProbsngluAlaArgGlnLeuSerLeuLeuArgAlaGlyAlaLeuIleAlaPro 432
Qy 224 CCCAGGTTCTCACTGAAGACATCTCTTCTGATCTTGGAAAAACAATGACACA 283
Db 433 ILeuInIleValGluGluArgThrIleGlyProThrLeuGlyMetGlnAsnIleGluIn 452
Qy 284 GGCATTATCTCAGCATGCTGGCTGGCAATGCTTATGTTTGGAGGATATATAT 343
Db 453 GylLeuGluAlaCysLeuAlaGlyLeuLeuValSerIleLeuPheMetIleIlePheTy 472
Qy 344 AGATTGGAGGCGTCATCGCTTGGAGCGTGTCTTCTGAATCTTTGCTTATCTGGCA 403
Db 473 LysIlePheGlyLeuIleAlaThrSerAlaLeuIleAlaAsnLeuIleLeuIleValGly 492
Qy 404 GCTCTACAGTATTG--GATGGCCACTCACTTGTAGAGACTGCTGGGATTTCTT 460
Db 493 ILeuSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512
Qy 461 GCTATGGGAGTGGCCGTGATGCAATGCTTGTATTCGAAGAATCGAGAGGATTT 520
Db 513 ThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGluArgIleGlyGluGluLeu 532
Qy 521 TTATTGTCTCAAGCTTAAATAATCTGTAGAAAAGATATACCAAGCTTTTGAGCC 580
Db 533 SerAsnGlyArgThrValGlnGlnAlaIleAspGluGlyTyArgGlyAlaPheSerSer 552
Qy 581 ATTTTGTCTCACTGACTACACTATGCGCTCAGCACTCTTTCTTCTTACATACA 640
Db 553 ILeuPheAspAlaAsnIleThrThrLeuIleValIleIleLeuValAlaGlyThr 572
Qy 641 GGGCTATTAAAGGTTGCTTGTGACATTTAGGAATTTCTCTCAAGTTTACG 700
Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaThrSerMetPheThr 592
Qy 701 GCTCTTTTCATGACTTAA 718
Db 593 AlaIleValGlyThrArg 598

RESULT 44
C90686
protein-export membrane protein secD [imported] - Escherichia coli (strain O157:H7, sub
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: C90686
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kunara, S.; Shiba, T.; Hatтори, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629, MUID:21156231, PMID:11258796
A:Accession: C90686
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-615 <HY>
A:Cross-references: GB:BA000007; PIDN:BA833882.1; PID:G13359916; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509552
C:Genetics:
A:Gene: EC60459
C:Superfamily: protein export membrane protein secD

Alignment Scores:
Pred. No.: 3.23e-23 Length: 615
Score: 359.50 Matches: 80
Percent Similarity: 62.39% Conservative: 61
Best Local Similarity: 35.40% Mismatches: 82
Query Match: 10.85% Indels: 3
DB: 2 Gaps: 2

US-09-868-987-1 (1-1864) x C90686 (1-615)
Qy 44 AATGACATATTTCTGCAACCGTGAAGCGATGCTGATGATTTAGCGTTATATG 103
Db 375 SerGlyLysLysAspAlaAsn-----GlyArgAlaValLeuValLysGlnGluGluVal 392

```

```

Qy 104 GTCAGACGCCCTATTTTAAACGTCCACTGAAATATGACGATGCTCAGGGAATTT 163
Db 393 ILeuAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleAsn 412
Qy 164 ACCACCGTGAAGTGAAGCAAACTCGCTCAGATTTTAAATCTGAGGAGATCTTTGTT 223
Db 413 AsnProbsngluAlaArgGlnLeuSerLeuLeuArgAlaGlyAlaLeuIleAlaPro 432
Qy 224 CCCAGGTTCTCACTGAAGACATCTCTTCTGATCTTGGAAAAACAATGACACA 283
Db 453 GylLeuGluAlaCysLeuAlaGlyLeuLeuValSerIleLeuPheMetIleIlePheTy 472
Qy 344 AGATTGGAGGCGTCATCGCTTGGAGCGTGTCTTCTGAATCTTTGCTTATCTGGCA 403
Db 473 LysIlePheGlyLeuIleAlaThrSerAlaLeuIleAlaAsnLeuIleLeuIleValGly 492
Qy 404 GCTCTACAGTATTG--GATGGCCACTCACTTGTAGAGACTGCTGGGATTTCTT 460
Db 493 ILeuSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512
Qy 461 GCTATGGGAGTGGCCGTGATGCAATGCTTGTATTCGAAGAATCGAGAGGATTT 520
Db 513 ThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGluArgIleGlyGluGluLeu 532
Qy 521 TTATTGTCTCAAGCTTAAATAATCTGTAGAAAAGATATACCAAGCTTTTGAGCC 580
Db 533 SerAsnGlyArgThrValGlnGlnAlaIleAspGluGlyTyArgGlyAlaPheSerSer 552
Qy 581 ATTTTGTCTCACTGACTACACTATGCGCTCAGCACTCTTTCTTCTTACATACA 640
Db 553 ILeuPheAspAlaAsnIleThrThrLeuIleValIleIleLeuValAlaGlyThr 572
Qy 641 GGGCTATTAAAGGTTGCTTGTGACATTTAGGAATTTCTCTCAAGTTTACG 700
Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaThrSerMetPheThr 592
Qy 701 GCTCTTTTCATGACTTAA 718
Db 593 AlaIleValGlyThrArg 598

RESULT 45
G85536
protein-export membrane protein secD [imported] - Escherichia coli (strain O157:H7, su
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Nov-2001
C:Accession: G85536
R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodac
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480, MUID:21074935, PMID:11206551
A:Accession: G85536
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-615 <STO>
A:Cross-references: GB:AE005174; NID:G12513258; PIDN:AA954755.1; GSPDB:GN00145; UMGP:Z
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: secD
C:Superfamily: protein export membrane protein secD

Alignment Scores:
Pred. No.: 3.23e-23 Length: 615
Score: 359.50 Matches: 80
Percent Similarity: 62.39% Conservative: 61
Best Local Similarity: 35.40% Mismatches: 82
Query Match: 10.85% Indels: 3
DB: 2 Gaps: 2

```

US-09-868-987-1 (1-1864) x G85536 (1-615)

```
Qy 44 AATGACAAATATCTGCAACCGTGGATGGCTGTAGTCATTGACCGTTATATG 103
Db 375 SerGlyLysLysAspAlaAsn-----GlyArgAlaValLeuValLysGlnGluVal 392
Qy 104 GTCAGCAGCCCTATTATTAACGTCCTCCATTGAAATCATGCGAGTGCTCAGGGAATTT 163
Db 393 IleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleAsn 412
Qy 164 ACCCACCCTGAAGTGAGCAAACTCGCTCAGATTAAATCTGGAGCGATGCTCTTTTGT 223
Db 413 AsnProAsnGluAlaArgGlnLeuSerLeuLeuLeuArgAlaGlyAlaLeuIleAlaPro 432
Qy 224 CCCGAGGTTCTCAGTGAGAGACGATCTCTTCATCTTGGGAAAAAACAATGTACACAA 283
Db 433 IleGlnIleValGluGluArgThrIleGlyProThrLeuGlyMetGlnAsnIleGluGln 452
Qy 284 GGCATTATCTCAGCATGCTGTGGCTTGCCAATGCTTATTTTGTAGAGCGGTATATAT 343
Db 453 GlyLeuGluAlaCysLeuAlaGlyLeuValSerIleLeuPheMetIleIlePheTyr 472
Qy 344 AGATTGGAGCGCTCATCGCTTCGGAGCTGTCTCTCTGAATCTTTTCTTATCTGGSCA 403
Db 473 LysLysPheGlyLeuIleAlaThrSerAlaLeuIleAlaAsnLeuIleValGly 492
Qy 404 GCTCTCAGTATTTG---GATCGGCCACTCACCTTGTTCAGGACTCGCTGGGATTGTTCT 460
Db 493 IleMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512
Qy 461 GCTATGGGATGGCGGTAGATGCAATGTTCTTGTATTTCGAAAGAAATCCGAGAGGAATTT 520
Db 513 ThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGluArgIleLysGluGluLeu 532
Qy 521 TTATTGTCTCAAGTCTTAAATAATCTAGAAAAAGGATATACCAGGCTTTTGGAGCC 580
Db 533 SerAsnGlyArgThrValGlnGlnAlaIleAspGluGlyTyrArgGlyAlaPheSerSer 552
Qy 581 ATTTTGTACTTAACCTGACTACAGTATTTGGCTCAGCAGCTTCTTTTCTCTAGATACA 640
Db 553 IlePheAspAlaAsnIleThrThrLeuIleLysValIleIleLeuTyrAlaValGlyThr 572
Qy 641 GGGCTATTAAAGGTTTGCTTGCATTTGATTTTGAATTTTCTCTCAATGTTTACG 700
Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaThrSerMetPheThr 592
Qy 701 GCTCTTTTCATGACTAAA 718
Db 593 AlaIleValGlyThrArg 598
```

Search completed: December 17, 2002, 12:40:18  
Job time : 91.5 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 06:28:50 ; Search time 64 Seconds  
(without alignments)  
8931.961 Million cell updates/sec

Title: US-09-868-987-1  
Perfect score: 1864  
Sequence: 1 atggacttcgcattatgtc.....atgacaaattcagataatgc 1864

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 892724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	116.4	6-2	6321	4	US-09-221-017B-311
2	99.6	5-3	2301	4	US-09-134-001C-852
3	84	4-5	2211	4	US-09-462-844-1
4	42.2	2-3	7218	1	US-08-232-463-14
5	39.6	2-1	1469	3	US-09-074-912-1
6	39.6	2-1	1469	4	US-09-230-136-1
C 7	38.8	2-1	298	4	US-09-071-035-147
C 8	38.8	2-1	441	4	US-09-071-035-145
9	37.2	2-0	3455	2	US-08-861-464-3
10	37.2	2-0	3455	2	US-08-396-001-3
11	37.2	2-0	3455	4	US-09-323-433A-3
12	37.2	2-0	8050	4	US-09-491-362-11
13	37.2	2-0	8050	4	US-09-874-562-11
14	36.2	1-9	499	4	US-09-641-638-52
15	36.2	1-9	499	4	US-09-641-638-572
16	35.8	1-9	246240	2	US-08-724-394A-20
17	35.8	1-9	246240	2	US-08-724-394A-21
18	35.8	1-9	246240	2	US-08-724-394A-22
19	35.4	1-9	19124	2	US-08-487-826B-13
C 20	34.8	1-9	1245	4	US-09-134-001C-2535
21	34.8	1-9	5529	3	US-08-869-696-1
22	34.4	1-8	13121	4	US-08-961-527-126
23	33.8	1-8	38844	4	US-09-734-675-3
24	33.4	1-8	3893	4	US-09-221-017B-604
C 25	33.2	1-8	1023	2	US-08-757-653-175
C 26	33.2	1-8	1023	2	US-08-823-516-78
C 27	33.2	1-8	1023	3	US-08-759-038-114

C	28	33.2	1-8	1023	3	US-08-758-314-114	Sequence 114, App
	29	33.2	1-8	2476	4	US-08-844-274-11	Sequence 11, Appl
	30	33.2	1-8	5194	4	US-08-844-274-16	Sequence 16, Appl
	31	33.2	1-8	5194	4	US-08-844-274-17	Sequence 17, Appl
C	32	33.2	1-8	5679	4	US-08-844-274-10	Sequence 10, Appl
	33	33.2	1-8	6448	4	US-08-844-274-15	Sequence 15, Appl
C	34	33.2	1-8	6723	4	US-08-844-274-13	Sequence 13, Appl
	35	33.2	1-8	6723	4	US-08-844-274-14	Sequence 14, Appl
	36	33.2	1-8	7560	4	US-08-844-274-20	Sequence 20, Appl
C	37	33.2	1-8	8898	4	US-08-961-527-69	Sequence 69, Appl
	38	33	1-8	2304	1	US-08-464-266-1	Sequence 1, Appl
C	39	33	1-8	2304	1	US-08-464-272-1	Sequence 1, Appl
	40	33	1-8	2304	4	US-08-464-514-1	Sequence 1, Appl
C	41	33	1-8	2304	4	US-08-486-403-1	Sequence 1, Appl
C	42	32.6	1-7	504	4	US-09-149-476-221	Sequence 221, App
	43	32.4	1-7	1260	3	US-09-008-979A-3	Sequence 3, Appl
	44	32.4	1-7	1260	4	US-09-460-618-3	Sequence 3, Appl
	45	32.4	1-7	1260	4	US-09-310-235B-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-09-221-017B-311/c  
; Sequence 311, Application US/09221017B  
; Patent No. 6444799  
; GENERAL INFORMATION:  
; APPLICANT: ROSS, Bruce C.  
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
; NUMBER OF SEQUENCES: 1120  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/221,017B  
; FILING DATE: 23-DEC-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP1182  
; FILING DATE: 31-DEC-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP1546  
; FILING DATE: 30-JAN-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP2911  
; FILING DATE: 09-APR-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/AU98/01023  
; FILING DATE: 10-DEC-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MONROY, Gladys H  
; REGISTRATION NUMBER: 32,430  
; REFERENCE/DOCKET NUMBER: 27340-20021.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-813-5600  
; TELEFAX: 650-494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 311:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6321 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)  
 ;  
 ; HYPOTHEICAL: NO  
 ; ANTI-SENSE: UNKNOWN  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: PORYPHYROMONAS GINGIVALIS  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: 1...6321  
 ;  
 ; US-09-221-017B-311

Query Match 6.2%; Score 116.4; DB 4; Length 6321;  
 Best Local Similarity 50.8%; Pred. No. 7.7e-22;  
 Matches 304; Conservative 0; Mismatches 291; Indels 3; Gaps 1;

QY 150 TCTCAGGAAATTTACCCAGCGTGAAGTGAAGAACTGCGCTCAGATTAAATCTGAG 209  
 DB 2009 TCTCCGGGACCTTACCCGTGAGAGGCGCGGTGACCTTGCCAGTAACTCCGGTA 1950  
 QY 210 CGATGCTTTTGTCCGAGGTTCTCAGTGAAGAGACATCTTCTGATCTTGGGAAA 269  
 DB 1949 AAATGAGATGCTACGCTAAAGCATCGAACAGAAAGTATGGTCTTAAGCTGGGTCGG 1890  
 QY 270 AACATGATACAAAGCATATATCAGCATGCTGCGCTTGGCAATGCTATTTGTTGA 329  
 DB 1889 AGTCATTAAAGCCGGATCTGTGCTCCGCTCGCTTGGTATCTGATGTGTACA 1830  
 QY 330 TGACCGTATATTA--TAGATTTGAGGCGTCAATGCTTGGGAGCTGTTCTTGAATC 386  
 DB 1829 TGCTGCGCTTACGGTTTCTTGCGGCGGTCTTAATCGCAAAAGCGGCATGATGTGAACA 1770  
 QY 387 TTTTGCTATCTGGGAGCTCTACAGTATTGGATGCGCCCATCACTTGCAGAGACTCG 446  
 DB 1769 GCTTCTTACATGCGCGTATTTGGCTTTTTCATGCGCGTGAACCTCTCGGGTATCG 1710  
 QY 447 CTGGGATTTGTTCTTGTCTATGAGGSAATGCGCTAGATGCAATGTTCTTGTATCGAAGAA 506  
 DB 1709 CAGTTTGTCTGACGCTGGGTATGCTGTGATGCCAAGTACTTATCTTCAGCGTA 1650  
 QY 507 TCCGAGAGAAATTTTATTTGTCTCAAGTCTTAAAAATCTGTAGAAAAAGATATACCA 566  
 DB 1449 TCAAGAAGAACTTCGTGCGGTGAAGACTCGAATTCGTGCGGTAAAGGATGTATGACA 1590  
 QY 567 AGGCTTTTGGAGCATTTTATTTCTATCTGACTAGAGTATGCGCTCAGACACTCTCTT 626  
 DB 1589 ACGCTTCTCTGCAATCTTCACTGAACTGTAACGACTATATTAACGGTATCAATCCAT 1530  
 QY 627 TCTTCTAGATACAGGCGCTATTAAGGGTTGCTTTGACATGATTTTAGAAATTTTCT 686  
 DB 1529 TCCCTACAGGAGCGGGGCGCATTCGCGGTTTGCACATGATGATTAATGCTCTATCG 1470  
 QY 687 CTTCATGTTTACGCGCTCTTTTCAATGACTAAATTTTCTTCAATGCTGTGATGAATAA 744  
 DB 1469 CTTCTTTCATTAAGCGCTGTCTTCTTGAATGCTTCTGAGAAACTGCGCAAAAA 1412

RESULT 2  
 ; US-09-134-001C-852  
 ; Sequence 852, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 852  
 ; LENGTH: 2301

TYPE: DNA  
 ; ORGANISM: Staphylococcus epidermidis  
 ; US-09-134-001C-852

Query Match 5.3%; Score 99.6; DB 4; Length 2301;  
 Best Local Similarity 48.7%; Pred. No. 1.8e-17;  
 Matches 270; Conservative 0; Mismatches 284; Indels 0; Gaps 0;

QY 173 GAAGTGAACAACCTCGCTCAGATTAAATCTGAGCCAGTCTTTGTTCCCGAGGTT 232  
 DB 694 GAAGCGAAACAAATAGCTGATTAATAATGCGGCTCATACAGTGTATTAATAAGAA 753  
 QY 233 CTCAGTAAAGAGACATCTCTTGTGATCTTGGGAAAAACAATGTACAGAGCATTAAC 292  
 DB 754 ATTTACTCTACCTGTGTGATGCAATTTGGTCAAGATGCTTTGATTAAGACATGTTT 813  
 QY 293 TCAGATGCTGTGGCTTGGCAATGCTTAATTTTGAATGAGCTATATTAATTAATTTGGA 352  
 DB 814 GCATCAATTTAGTATAGATTAATTAATTTATGCTTGTCTATCGTTTGCTT 873  
 QY 353 GCGGTCACTGCTTGGGAGCTGTTCTTGAATCTTTTGTCTATCTGGGAGCTTACAG 412  
 DB 874 GGTATGTTGCAATCAATGCTTAAACCATTAATTAATTAATTAATTAATTAATTAAT 933  
 QY 413 TATTTGATGCGCACTCACTGTGAGAGCTGCGGATGTTGTTGCTATGGGATG 472  
 DB 934 TTTATATCAGGTATTAATCTTAATCTTAATCTTAATCTTAATCTTAATCTTAATCT 993  
 QY 473 GCGGTAGATCAATATGTTCTTGTATTCGAAAGATCCGAGAGAAATTTTATGTTCTCAA 532  
 DB 994 GCTGTGATCCCAATATCAATATGATGAACGATTAATTAATTAATTAATTAATTAAT 1053  
 QY 533 AGCTTAAAAAATCTGTAGAAAAAGATATCAAGGCTTTTGGAGCATTTTGAATCT 592  
 DB 1054 ACCCTTAAACAAAGCTATTAACAAAGCAATTAAGTCAATCTTAATTAATTAATTAAT 1113  
 QY 593 AACTGATCTACATTTGGCTCAGACCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 652  
 DB 1114 AACTTAACAATCTGATGCTGAGCTGTGCTTTTCTTCTTCTTCTTCTTCTTCTTCTT 1173  
 QY 653 GGGTTTCTTGTGATTTTGAATTTTGAATTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 712  
 DB 1174 GCGTTCGAACCAATGTTACTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTT 1233  
 QY 713 ACTTAATTTTCTT 726  
 DB 1234 TCAAGAGGGTGT 1247

RESULT 3  
 ; US-09-462-844-1  
 ; Sequence 1, Application US/09462844  
 ; Patent No. 6258563  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Quax, Wilhelmus J.  
 ; TITLE OF INVENTION: Increasing Production of Proteins in  
 ; TITLE OF INVENTION: Gram-Positive Microorganisms  
 ; FILE REFERENCE: GC385-US  
 ; CURRENT APPLICATION NUMBER: US/09/462,844  
 ; CURRENT FILING DATE: 2000-01-13  
 ; PRIOR APPLICATION NUMBER: PCT/US98/14786  
 ; PRIOR FILING DATE: 1998-07-15  
 ; PRIOR APPLICATION NUMBER: EP 97305286.3  
 ; PRIOR FILING DATE: 1997-07-16  
 ; PRIOR APPLICATION NUMBER: EP 97305344.0  
 ; PRIOR FILING DATE: 1997-07-17  
 ; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 2211  
 ; TYPE: DNA  
 ; ORGANISM: Bacillus subtilis  
 ; US-09-462-844-1





```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,912
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: SPRUILL, W. MURRAY
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: 5718-9
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919 420 2202
TELEFAX: 919 881 3175
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1469 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Pentaclethra macroloba
FEATURE:
NAME/KEY: CDS
LOCATION: 31..1257
US-09-074-912-1

Query Match      2.1%; Score 39.6; DB 3; Length 1469;
Best Local Similarity 50.5%; Pred. No. 0.28; Indels 0; Gaps 0;
Matches 96; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 1315 TGAATGCAATATGCTTTCAGTGCCTTATGCGCTTATTCATGACCTTTGGCTACCTG 1374
    |||||
DB 678 TGAATTCATCTCGTTGATGTGCAATCATGCTGATTTCCGCCCGGTGCTCTCAG 737

QY 1375 TGCAGTCTTGTTATAGCACATTTCTTTGAAGAAATTAATTAATTTGCAAGCCAT 1434
    |||||
DB 738 CGAGTGCTCCAGCAAGAAATACAGAAATTAAGAAATCTTTGCTGCTATATAGGAAC 797

QY 1435 TGGTGCCTTATGACTGATGTTGGGGTATTCATTAACAATCCTTGAATTTGATCG 1494
    |||||
DB 798 TGGAGTTGTAAACCTGGTGAGGGTTATTCGTTATGCTGACTGACTATTTTCGATTG 857

QY 1495 TATTCGTGAA 1504
    |||||
DB 858 GAGTAGTGAA 867

RESULT 6
US-09-290-136-1
Sequence 1, Application US/09290136
Patent No. 6339144
GENERAL INFORMATION:
APPLICANT: Cigan, Amy L.
APPLICANT: Czaplá, Thomas H.
APPLICANT: Fallis, Lynn
APPLICANT: Meyer, Terry E.
APPLICANT: Mundell, Scott A.
APPLICANT: Sabus, Brian
APPLICANT: Schubert, Karel
TITLE OF INVENTION: Proteins Having Insecticidal Activities and Methods of
TITLE OF INVENTION: Use
FILE REFERENCE: 5718-9A, 035718/180486
CURRENT APPLICATION NUMBER: US/09/290,136
CURRENT FILING DATE: 1999-04-13
EARLIER APPLICATION NUMBER: 09/074,912
EARLIER FILING DATE: 1998-05-08
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
```

```
LENGTH: 1469
TYPE: DNA
ORGANISM: Pentaclethra macroloba
FEATURE:
NAME/KEY: CDS
LOCATION: (31)..(1257)
US-09-290-136-1

Query Match      2.1%; Score 39.6; DB 4; Length 1469;
Best Local Similarity 50.5%; Pred. No. 0.28; Indels 0; Gaps 0;
Matches 96; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 1315 TGAATGCAATATGCTTTCAGTGCCTTATGCGCTTATTCATGACCTTTGGCTACCTG 1374
    |||||
DB 678 TGAATTCATCTCGTTGATGTGCAATCATGCTGATTTCCGCCCGGTGCTCTCAG 737

QY 1375 TGCAGTCTTGTTATAGCACATTTCTTTGAAGAAATTAATTAATTTGCAAGCCAT 1434
    |||||
DB 738 CGAGTGCTCCAGCAAGAAATACAGAAATTAAGAAATCTTTGCTGCTATATAGGAAC 797

QY 1435 TGGTGCCTTATGACTGATGTTGGGGTATTCATTAACAATCCTTGAATTTGATCG 1494
    |||||
DB 798 TGGAGTTGTAAACCTGGTGAGGGTTATTCGTTATGCTGACTGACTATTTTCGATTG 857

QY 1495 TATTCGTGAA 1504
    |||||
DB 858 GAGTAGTGAA 867

RESULT 7
US-09-071-035-147/c
Sequence 147, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44M storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brooks
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 147:
SEQUENCE CHARACTERISTICS:
LENGTH: 298 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-147

Query Match      2.1%; Score 38.8; DB 4; Length 298;
Best Local Similarity 50.0%; Pred. No. 0.23;
```

```

QY 1713 CACCTCTGTTGTTGTTATGTCGCTAAAGAAAAATCGCTCAAAAATAAGTACCGTTAAACT 1772
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 277 TATCACTAAATTTTGAATTATACCTCGAAAAAGTTTCATCATCAAAAAAAGTGTAAGTAG 218
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1773 TAATCTAAGCTGTA 1786
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 217 TAGTTGGACTTTTA 204
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-08-861-464-3
; Sequence 3, Application US/08861464
; Patent No. 5874210
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: Genes Determining Cellular Senescence
; TITLE OF INVENTION: in Yeast
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/861,464
; FILING DATE: 22-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/396,001
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/09351
; FILING DATE: 15-AUG-1994
; APPLICATION NUMBER: US 08/107,408
; FILING DATE: 16-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6408A22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 781-861-6240
; TELEFAX: 781-861-9540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3455 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 663..3164
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 663..3164
;
US-08-861-464-3

Query Match 2.0%; Score 37.2; DB 2; Length 3455;
Best Local Similarity 57.9%; Pred. No.1.8;
Matches 66; Conservative 0; Mismatches 48; Indels 0; Gaps 0
QY 522 TATTGTCTCAAAAGTCTTAAAAAATCTGTAGAAAAAGGATATACCAAGCGCTTTTGGAGCCA 581
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 2207 TAGTAACGCAATTGGGCAAAACAGCTCTAGTACATTAAGTAATTAAGCA 2266  
Qy 582 TTTTGAATTCATTAAGTACATTAAGTATTTGGCTCAGCACTTTCTTCTAG 635  
Db 2267 TTTTCAATGAAATGGTAACTTGTCTTCAAAAGAAATTAATTAAGCA 2320

RESULT 10  
US-08-396-001-3  
Sequence 3, Application US/08396001  
Patent No. 5919618

## GENERAL INFORMATION:

APPLICANT: Guarente, Leonard P.  
APPLICANT: Austriaco Jr., Nicanor  
APPLICANT: Claus, James  
APPLICANT: Cole, Francesca  
APPLICANT: Kennedy, Brian  
TITLE OF INVENTION: Genes Determining Cellular Senescence in  
TITLE OF INVENTION: Yeast  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/396,001  
FILING DATE: 28-FEB-1995  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: MIT-6408A2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540

## INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 3455 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

FEATURE:  
NAME/KEY: CDS  
LOCATION: 663..3164  
US-08-396-001-3

Query Match 2.0%; Score 37.2; DB 2; Length 3455;  
Best Local Similarity 57.9%; Pred. No. 1.8;  
Matches 66; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 522 TATGTCTCAAGCTTAAATAATCTGAGAAAAGATATACCAAGCTTTGGAGCCA 581  
Db 2207 TAGTAACGCAATTGGGCAAAACAGCTCTAGTACATTAAGTAATTAAGCA 2266  
Qy 582 TTTTGAATTCATTAAGTACATTAAGTATTTGGCTCAGCACTTTCTTCTAG 635  
Db 2267 TTTTCAATGAAATGGTAACTTGTCTTCAAAAGAAATTAATTAAGCA 2320

RESULT 11  
US-09-323-433A-3  
Sequence 3, Application US/09323433A  
Patent No. 6218512  
GENERAL INFORMATION:

APPLICANT: Guarente, Leonard P.  
APPLICANT: Austriaco Jr., Nicanor  
APPLICANT: Claus, James J.  
APPLICANT: Cole, Francesca

TITLE OF INVENTION: GENES DETERMINING CELLULAR SENESCENCE IN  
TITLE OF INVENTION: YEAST  
FILE REFERENCE: 0050,1491-003  
CURRENT APPLICATION NUMBER: US/09/323,433A  
CURRENT FILING DATE: 1999-06-01  
PRIOR APPLICATION NUMBER: US 08/396,001  
PRIOR FILING DATE: 1995-02-28  
PRIOR APPLICATION NUMBER: PCT/US94/09351  
PRIOR FILING DATE: 1994-08-15  
PRIOR APPLICATION NUMBER: US 08/107,408  
PRIOR FILING DATE: 1993-08-16

NUMBER OF SEQ ID NOS: 48  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3

LENGTH: 3455

TYPE: DNA

ORGANISM: Saccharomyces cerevisiae

FEATURE:  
NAME/KEY: CDS  
LOCATION: (663)...(3164)

OTHER INFORMATION: UTH4  
US-09-323-433A-3

Query Match 2.0%; Score 37.2; DB 4; Length 3455;  
Best Local Similarity 57.9%; Pred. No. 1.8;  
Matches 66; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 522 TATGTCTCAAGCTTAAATAATCTGAGAAAAGATATACCAAGCTTTGGAGCCA 581  
Db 2207 TAGTAACGCAATTGGGCAAAACAGCTCTAGTACATTAAGTAATTAAGCA 2266  
Qy 582 TTTTGAATTCATTAAGTACATTAAGTATTTGGCTCAGCACTTTCTTCTAG 635  
Db 2267 TTTTCAATGAAATGGTAACTTGTCTTCAAAAGAAATTAATTAAGCA 2320

## RESULT 12

US-09-491-362-11  
Sequence 11, Application US/09491362  
Patent No. 6281017

## GENERAL INFORMATION:

APPLICANT: Croteau, Rodney B  
APPLICANT: Lange, Bernd M  
TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND  
TITLE OF INVENTION: METHODS OF USE  
FILE REFERENCE: MSUR14977  
CURRENT APPLICATION NUMBER: US/09/491,362  
CURRENT FILING DATE: 2000-01-26  
EARLIER APPLICATION NUMBER: 60/118,349  
EARLIER FILING DATE: 1999-02-03  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 11

LENGTH: 8050

TYPE: DNA

ORGANISM: Arabidopsis thaliana  
US-09-491-362-11

Query Match 2.0%; Score 37.2; DB 4; Length 8050;  
Best Local Similarity 53.4%; Pred. No. 2.7;  
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 1363 TTTGCTACCTGTCAGCTGTTTATAGACAAATTTCTTTGAAGAAATTCAAATAGA 1422  
Db 7423 TTTGCTCCGCTTCGCTTTTGTACATAGTAGATTTTATATAGAGCTGCTAATGAGGT 7482  
Qy 1423 TTTGCAAGCAATGGTCTTAAATGAATTAAGTATTTGGGATATTCATTAACAATCTTGAT 1482

```

RESULT 14
US-09-641-638-52
; Sequence 52, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
; FILE REFERENCE: GENSET.051CPI
; CURRENT APPLICATION NUMBER: US/09/641.638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 52

```

LENGTH: 499  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: allele  
LOCATION: 271  
OTHER INFORMATION: 12-215-272 : deletion T  
NAME/KEY: misc\_binding  
LOCATION: 251..270  
OTHER INFORMATION: 12-215-272.misl, potential  
NAME/KEY: primer\_bind  
LOCATION: 1..20  
OTHER INFORMATION: upstream amplification primer  
NAME/KEY: primer\_bind  
LOCATION: 479..499  
OTHER INFORMATION: downstream amplification primer, complement  
NAME/KEY: misc\_feature  
LOCATION: 87  
OTHER INFORMATION: n=a, g, c or t  
US-09-641-638-572

Query Match  
Best Local Similarity 1.9%; Score 36.2; DB 4; Length 499;  
Best Local Similarity 50.2%; Pred. No. 1.5;  
Matches 115; Conservative 0; Mismatches 113; Indels 1; Gaps 1;

QY 1322 CAAATGCTTTCAGTCCGATGCGCTTTAATTCATGACCTTTGGCTACCTGTCAGTC 1381  
DB 217 CATTGAGTTCAGTCTTTGATGCTATGTTAAATAGGTCTAATTTCTTTTAAATGT 276  
QY 1382 TTGTTAATGACATTTCTTTTGAAGAAATTCAAATAGAT-TTGAAGCATTTGTC 1440  
DB 277 TTAAGTCAGCTCTTTGTTTGTCTCAATTTCTTTAGATTAAGAACTTTTGTAG 336  
QY 1441 TTTAATGCTGATTTGGGTATTCATTAACATATCTTGATTCATTTTGAATCGTATTCG 1500  
DB 337 CTCATATACAGTTTAGATTTCTTTTTCATTTTCAGTGAAGAAATGCAATTTTGA 396  
QY 1501 TGAAGATCCCAAGCAACCTGTTTACCCTATGATGATTTTATTTAT 1549  
DB 397 GAAAGGGGTTCAGCGAATCTTTAGATCATTGGGTAGTAGACATT 445

RESULT 16  
US-08-724-394A-20  
Sequence 20, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Kronmal, Gregory S.  
APPLICANT: Lauer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchihashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
TITLE OF INVENTION: Sequences and Antibodies Thereo  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:  
NAME: Flits, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ. ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note="HLA-H.CONTIG"  
US-08-724-394A-20

Query Match  
Best Local Similarity 1.9%; Score 35.8; DB 2; Length 246240;  
Best Local Similarity 52.3%; Pred. No. 28;  
Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 1348 TTTAATTCATGACCTTTTGGCTACCTGTGCGAGCTTTATATGACATTTCTTTTGA 1407  
DB 85039 TTTAAGCTTCAAAATTTCTTTAAATTTCTGCTTAGCTTCAATTTTATATAA 85098  
QY 1408 GAAATTCATATGATTTTGGACCATTTGGCTTTAATGACTGTATTTGGGTATTCAT 1467  
DB 85099 TAACTTTTATATATGTTTCATTTCTATACATCTTCTAATGACTTATATAGTTTACAT 85158  
QY 1468 AAACATATCTTGATCATTTTGGATGATTT 1498  
DB 85159 TCCATATATTTAGATCTTTTATGATCATCT 85189

RESULT 17  
US-08-724-394A-21  
Sequence 21, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Kronmal, Gregory S.  
APPLICANT: Lauer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchihashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
TITLE OF INVENTION: Sequences and Antibodies Thereo  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Flits, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H. CONTIG"  
US-08-724-394A-21

Query Match 1.9%; Score 35.8; DB 2; Length 246240;  
Best Local Similarity 52.3%; Pred. No. 28;  
Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
QY 1348 TTTAATTCATGACCTTTTGGCTGACCTGTCAGTCTTGTATAGCACATTTCTTTTGA 1407  
DB 85039 TTTAAGCGCTACAATTTCTCTTAAATTCGCTTAGCTTCATTTGATATAATTTTATA 85098  
QY 1408 GAAATTCGAATAGATTTGCAAGCCATTTGGTCTTAAATGACTGATTTGGGGTATTCA 1467  
DB 85099 TTAACCTTTTATTATTGTTCCATCTCTATACATCTCTCTAATGACTTATTAGTAGTTTACATT 85158  
QY 1468 AAACAATCTTGGATCATTTTGGATCGTATT 1498  
DB 85159 TCCTAATATTAGAACTCTTTTAGTCATCTT 85189

## RESULT 18

US-08-724-394A-22  
Sequence 22, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Kronmal, Gregory S.  
APPLICANT: Lauer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchihashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
TITLE OF INVENTION: Sequences and Antibodies Thereto  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitts, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs

TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H. CONTIG"  
US-08-724-394A-22

Query Match 1.9%; Score 35.8; DB 2; Length 246240;  
Best Local Similarity 52.3%; Pred. No. 28;  
Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
QY 1348 TTTAATTCATGACCTTTTGGCTGACCTGTCAGTCTTGTATAGCACATTTCTTTTGA 1407  
DB 85039 TTTAAGCGCTACAATTTCTCTTAAATTCGCTTAGCTTCATTTGATATAATTTTATA 85098  
QY 1408 GAAATTCGAATAGATTTGCAAGCCATTTGGTCTTAAATGACTGATTTGGGGTATTCA 1467  
DB 85099 TTAACCTTTTATTATTGTTCCATCTCTATACATCTCTCTAATGACTTATTAGTAGTTTACATT 85158  
QY 1468 AAACAATCTTGGATCATTTTGGATCGTATT 1498  
DB 85159 TCCTAATATTAGAACTCTTTTAGTCATCTT 85189

## RESULT 19

US-08-487-826B-13  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellem, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-9550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

US-08-487-826B-13

Query Match 1.9%; Score 35.4; DB 2; Length 19124;  
Best Local Similarity 44.7%; Pred. No. 12;  
Matches 181; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 1387 TATAGACATTTCTTTGGAAGAAATTCAAATGATTTGCAACCCATTTGCTTAAAT 1446  
DB 3560 TATATATTTTTTTTAACTGGATGCTTACATTTTATTTATATATATATTTAAAT 3619  
QY 1447 GACTGATTTGGGATTCATTTAAACAATCTTGATCATTTTGTATTCGTGAAGA 1506  
DB 3620 TATATATCTTATATATATATATATTTTACCTTAAGATATATATATATATATATC 3679  
QY 1507 TCGCCAAAGCAACCTGTTTACCCCTATGATGTTTGTATATATATGCCCCCTCAAAAGAC 1566  
DB 3680 TATCTATCTAT 3739  
QY 1567 GTTCAGCCGACCGTATATGACAGCTACACTCTATCAGTTTGTATATGCTTTGTT 1626  
DB 3740 ATTAGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3799  
QY 1627 TATAGCGGCTCTCTGCTTATTTGATTTATATGACATAGGATTCCTTAGG 1686  
DB 3800 TATAGAACTTCATTTTATTTATTTATATATATATATATATATATATATATATAT 3856  
QY 1687 AACTTATCGTCTTTATATATGACACACTCTGTTGTTTATGCTCGTAAAGAAA 1746  
DB 3857 AATTACACATGAT 3916  
QY 1747 TCGCTCAAAATTAATACCGTTAACTTAATCTAATGCTTACAGT 1791  
DB 3917 TTGTTTAAATATAGTTGTTATATATATATATATATATATATATATATATATAT 3961

## RESULT 20

US-09-134-001C-2535/C  
Sequence 2535, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 2535  
LENGTH: 1245  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2535

Query Match 1.9%; Score 34.8; DB 4; Length 1245;  
Best Local Similarity 51.3%; Pred. No. 5.3;  
Matches 81; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
QY 569 GCTTTTGAGCCATTTTGGATTTCTAATCTGACTACGATTTGGCTTCGACCTCTTTTC 628  
DB 311 GCAGGTGAACAGTGTTGATTACTTATCTGCTGTAATGTTTCTTGAATGTTTCA 252  
QY 629 TTCTAGATACAGGCTTATTAAGGTTGCTTGAATGATTTTGAATTTTCTCT 688  
DB 251 TTATATCTTGTAGTTTATTAACCTCGTTAGAAATTAATTAATTAATTAATTAATTA 192  
QY 689 TCAATGTTTACGGCTCTTTGATGATGATTAATTTTCTT 726  
DB 191 TCTATATTGCTCTTGAATTTAAGCTTGATGATTTCTT 154

## RESULT 21

US-08-869-696-1  
Sequence 1, Application US/08869696C  
Patent No. 6031155  
GENERAL INFORMATION:  
APPLICANT: Cameron-Mills, Verena  
APPLICANT: Lok, Finn  
APPLICANT: Sinjorgo, Catharina Maria Cornelia  
APPLICANT: Van Den Doel, Ronald Tako Marinus  
APPLICANT: Caspers, Martinus Petrus Maria  
APPLICANT: Van Zeijl-Van Der Valk, Maria Joanna  
TITLE OF INVENTION: ARABINOXYLAN DEGRADATION  
FILE REFERENCE: 11225.01US01  
CURRENT APPLICATION NUMBER: US/08/869,696C  
CURRENT FILING DATE: 1997-06-05  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 5529  
TYPE: DNA  
ORGANISM: barley  
US-08-869-696-1

Query Match 1.9%; Score 34.8; DB 3; Length 5529;  
Best Local Similarity 58.8%; Pred. No. 10;  
Matches 60; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1351 AATTCAGACCTTTGGCTACCTGTCAGCTTTGTTATAGCAATTTCTTTGAAGAA 1410  
DB 4320 AATTCAGAAATATTTCTTACTCATCAAGTTTGAATGATGACATTTTTCACAA 4379  
QY 1411 AATTCAGATGATTTGCAAGCCATTTGCTTAAAGACTGT 1452  
DB 4380 ATTCGAAATTAATTTTAAACATGTTTAAATTAATGAT 4421

## RESULT 22

US-08-961-527-126  
Sequence 126, Application US/08961527  
Patent No. 6420135  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 126:  
SEQUENCE CHARACTERISTICS:



LENGTH: 13121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-126

Query Match 1.8%; Score 34.4; DB 4; Length 13121;  
Best Local Similarity 52.0%; Pred. No. 19;  
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
Qy 589 TTCTAACTTGACATACAGATTTGGCTCAGCACTTCTTTTCTCTCTAGATACAGGCGCTAT 648  
Db 12944 TCCTAAGTGCACACTTGTATCCGCTCTGACGGCTTTTATGACACATTTATTTGGAT 13003  
Qy 649 TAAAGGTTTGTCTTGACATGATTTTAAAGAAATTTCTCTTCAATGTTTACGGCTTTTT 708  
Db 13004 CAATTGAATATCTTCGTAATCAAAATTTGAAATTCATTTAACAATATCGATGTCGTTT 13063  
Qy 709 CATGACTAAATTTTCTTCATGCTGTGG 736  
Db 13064 CTTTGTGTAATGACCTACCTATGCTCTTG 13091

## RESULT 23

US-09-734-675-3  
Sequence 3, Application US/09734675  
Patent No. 6365391  
GENERAL INFORMATION:  
APPLICANT: WEBSTER, Marion et al  
TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
TITLE OF INVENTION: USES THEREOF  
FILE REFERENCE: CL000862  
CURRENT APPLICATION NUMBER: US/09/734.675  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 38844  
TYPE: DNA  
ORGANISM: Human  
US-09-734-675-3

Query Match 1.8%; Score 33.8; DB 4; Length 38844;  
Best Local Similarity 48.2%; Pred. No. 44;  
Matches 124; Conservative 0; Mismatches 132; Indels 1; Gaps 1;  
Qy 501 AAAGAATCCGAGAGGAATTTTATTTGTTCTCAAGTCTTAAAAATCTGAGAAAAGGAT 560  
Db 5655 ACAGAAGCAGAGACTTATTTTAAAAAGAAAGAAATGACAAATGTTATTTGTTTAAAT 5714  
Qy 561 ATACCAAGGCTTTTGGAGCCATTTTGTATTTCTTAAGTCTGACTACAGTATTTGGCTCAGCAC 620  
Db 5715 CAAGAACCATTTTTAAGACACTTCTTTCCCAATCATCTACCAATTTTCTCTGTGATC 5774  
Qy 621 TCTTTTCTTCTAGATACAGGCGCTATTAAGGGTTTGTCTTTGACATGATTTTATGAAA 680  
Db 5775 ATTGCTCTTTGTCCTAGTATACCTAATGCGCATCATATTTACATAAT-ATTGTAGAGT 5833  
Qy 681 TTTTCTCTCATGTTTACGGCTCTTTTCATGACTAAATTTTCTTCTCATGCTGGATGA 740  
Db 5834 TTATAATCTCTATTTTTCAGTTAAACATTAAATCAATTCACAAATTTCTTAATTTTGTGGTTTC 5893  
Qy 741 ATAAGACCCCAACATACA 757  
Db 5894 ATCTTTCCCAACCAATA 5910

## RESULT 24

US-09-221-017B-604  
Sequence 604, Application US/09221017B  
Patent No. 6444799  
GENERAL INFORMATION:

APPLICANT: Ross, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PPI182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PPI546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Monroy, Gladys H  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 604:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3893 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORPHYROMONAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1...3893  
US-09-221-017B-604

Query Match 1.8%; Score 33.4; DB 4; Length 3893;  
Best Local Similarity 53.4%; Pred. No. 21;  
Matches 70; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
Qy 497 TTCGAAAGATCCGAGAGGAATTTTATTTGTTCTCAAGTCTTAAAAATCTGTAGAAAAA 556  
Db 2020 TCCGAAACAATGAGCTAAATCTCTTTCTGCATCACATAGGGCAATATCTTCAGATAA 2079  
Qy 557 GGATATACCAAGGCTTTTGGAGCCATTTTGTATCTTACTTGTACAGTATTCGCTCA 616  
Db 2080 AGTTGTCTGAGCTTTTGTATCTGTCTTGTAGAGCATAAAGAGATTTATGTTCTTAA 2139  
Qy 617 GCACTTCTTTT 627  
Db 2140 AGTCTTCAGTT 2150

## RESULT 25

US-08-757-653-175/c  
Sequence 175, Application US/08757653  
Patent No. 584369  
GENERAL INFORMATION:  
APPLICANT: Kaiser, Michael W.  
APPLICANT: Lyamichev, Victor I.  
APPLICANT: Lyamichev, Natasha  
TITLE OF INVENTION: Cleavage Of Nucleic Acid Using  
TITLE OF INVENTION: Thermostable FEN-1 Endonucleases  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/757,653  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: FORS-02565  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 175:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1023 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1020  
US-08-757-653-175

Query Match 1.8%; Score 33.2; DB 2; Length 1023;  
Best Local Similarity 49.4%; Pred. No. 13;  
Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 336 TATATTATGATTGGAGCGTCATCGCTTGGAGCTGTTCTTGAATCTTTTGTCTTA 395  
|||  
DB 825 TATTGCATATTAATCCACATCGCTTGTCTTTGGAACTTGTGCGGATCTTTTGAGTG 766  
|||  
QY 396 TCTGGGAGCTCTACAGATTGGATGGCCACCTGTCAGAGACTCGCTGGGATTG 455  
|||  
DB 765 TCTAACATCTCTAAGCTTTTAAAGCTTATGCGCTTATCTCTCTGGGTTGATGC 706  
|||  
QY 456 TTCTGTATGAGGATGGCCGTAGATGCAAAATGTTCTTGTAATTGAAAGATCC 509  
|||  
DB 705 TGTTCACAGAGATTGCTAGTCAATGAGCTTTCTCTGTAGACTTAATTC 652  
|||

RESULT 26  
US-08-823-516-78/c  
Sequence 78, Application US/08823516  
Patent No. 5994069  
GENERAL INFORMATION:  
APPLICANT: Hall, Jeff G.  
APPLICANT: Lyamichev, Victor I.  
APPLICANT: Mast, Andrea L.  
APPLICANT: Brown, Mary Ann D.  
TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple  
TITLE OF INVENTION: Sequential Invasive Cleavages

NUMBER OF SEQUENCES: 163  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/823,516  
FILING DATE: 24-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/01072  
FILING DATE: 21-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/759,038  
FILING DATE: 02-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/758,314  
FILING DATE: 02-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/756,386  
FILING DATE: 29-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/682,853  
FILING DATE: 12-JUL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/599,491  
FILING DATE: 24-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: FORS-02736  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 78:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1023 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1020  
US-08-823-516-78

Query Match 1.8%; Score 33.2; DB 2; Length 1023;  
Best Local Similarity 49.4%; Pred. No. 13;  
Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 336 TATATTATGATTGGAGCGTCATCGCTTGGAGCTGTTCTTGAATCTTTTGTCTTA 395  
|||  
DB 825 TATTGCATATTAATCCACATCGCTTGTCTTTGGAACTTGTGCGGATCTTTTGAGTG 766  
|||  
QY 396 TCTGGGAGCTCTACAGATTGGATGGCCACCTGTCAGAGACTCGCTGGGATTG 455  
|||  
DB 765 TCTAACATCTCTAAGCTTTTAAAGCTTATGCGCTTATCTCTCTGGGTTGATGC 706  
|||  
QY 456 TTCTGTATGAGGATGGCCGTAGATGCAAAATGTTCTTGTAATTGAAAGATCC 509  
|||  
DB 705 TGTTCACAGAGATTGCTAGTCAATGAGCTTTCTCTGTAGACTTAATTC 652  
|||

RESULT 27  
US-08-759-038-114/c

; Sequence 114, Application US/08759038  
; Patent No. 6090543  
; GENERAL INFORMATION:  
; APPLICANT: Prudent, James R.  
; APPLICANT: Hall, Jeff G.  
; APPLICANT: Lyamichev, Victor I.  
; APPLICANT: Brow, Mary Ann D.  
; APPLICANT: Dahlberg, James E.  
; TITLE OF INVENTION: Cleavage Of Nucleic Acids  
; NUMBER OF SEQUENCES: 134  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States Of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/759,038  
; FILING DATE: 02-DEC-1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/  
; FILING DATE: 29-NOV-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/682,853  
; FILING DATE: 12-JUL-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/599,491  
; FILING DATE: 24-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: FORS-02574  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 114:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1023 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1020  
; US-08-759-038-114

Query Match 1.8%; Score 33.2; DB 3; Length 1023;  
Best Local Similarity 49.4%; Pred. No. 13;  
Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;  
Qy 336 TATATTATAGATTGGAGCGTCATCGCTTCGGAGCTGTTCTTCTGAATCTTTTGCTTA 395  
Db 825 TATTGCATATAAATCCATCGCTTTCCTTTTGGAACTTGTCTAGCGGATCTTTGAGTG 766  
Qy 396 TCTGGGAGCTCTACAGTATTTGGATGCGCCACTCACCTTGTGACGAGCTCGCTGGGATTG 455  
Db 765 TCTAACATCTCTAAAGCTTTTAAAGCCCTATGCCCTTTATCTCTCTGGTGTAGTC 706  
Qy 456 TTCTTGCTATGGGATGCGCGTAGATGCAAAATGTTCTTGTATTCGAAAGATCC 509  
Db 705 TGTTCACACGAGGATTGCTAGTTCATAGCTTTTCTCTTGTAGCTTTAATTC 652

RESULT 28  
US-08-758-314-114/c

; Sequence 114, Application US/08758314  
; Patent No. 6090606  
; GENERAL INFORMATION:  
; APPLICANT: Kaiser, Michael W.  
; APPLICANT: Lyamichev, Victor I.  
; APPLICANT: Lyamichev, Natasha  
; TITLE OF INVENTION: Improved Cleavage Agents  
; NUMBER OF SEQUENCES: 134  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States Of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/758,314  
; FILING DATE: 02-DEC-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/  
; FILING DATE: 29-NOV-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/682,853  
; FILING DATE: 12-JUL-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/599,491  
; FILING DATE: 24-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: FORS-02575  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 114:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1023 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1020  
; US-08-758-314-114

Query Match 1.8%; Score 33.2; DB 3; Length 1023;  
Best Local Similarity 49.4%; Pred. No. 13;  
Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;  
Qy 336 TATATTATAGATTGGAGCGTCATCGCTTCGGAGCTGTTCTTCTGAATCTTTTGCTTA 395  
Db 825 TATTGCATATAAATCCATCGCTTTCCTTTTGGAACTTGTCTAGCGGATCTTTGAGTG 766  
Qy 396 TCTGGGAGCTCTACAGTATTTGGATGCGCCACTCACCTTGTGACGAGCTCGCTGGGATTG 455  
Db 765 TCTAACATCTCTAAAGCTTTTAAAGCCCTATGCCCTTTATCTCTCTGGTGTAGTC 706  
Qy 456 TTCTTGCTATGGGATGCGCGTAGATGCAAAATGTTCTTGTATTCGAAAGATCC 509  
Db 705 TGTTCACACGAGGATTGCTAGTTCATAGCTTTTCTCTTGTAGCTTTAATTC 652

RESULT 29  
US-08-844-274-11  
; Sequence 11, Application US/08844274B  
; Patent No. 6218185

Query Match	1.8%;	Score 33.2;	DB 4;	Length 5194;
Best Local Similarity	46.5%;	Pred. No. 27;		

RESULT 32  
US-08-844-274-10/C  
; Sequence 10, Application US/08844274E  
; Patent No. 6218185  
; GENERAL INFORMATION:  
; APPLICANT: Frazer Jr., Malcom J.





```

; ADDRESSES: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: United States
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,272
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/013,975
; FILING DATE: 04-FEB-1993
; APPLICATION NUMBER: US 07/497,935
; FILING DATE: 22-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9350
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2304 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 163..1704
;
; US-08-464-272-1
;
; Query Match 1.8%; Score 33; DB 1; Length 2304;
; Best Local Similarity 63.0%; Pred.No.21;
; Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps
;
; QY 646 TATTAAGGGTTGCTTTGACATTCATTAGGAATTTCTCTCAATGTTTACGGCTCT 705
; ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 2009 TAGTTTAACTTTACTTTTACTTTGCTGTGCGATTTTCGTTTGTGTCT 1950
; ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
;
; QY 706 TTTCATGACTAAATTTTCTT 726
; ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 1949 TTTCAGATTTTCGATTTTGT 1929
; ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
;
; RESULT 40
; US-08-464-514-1/c
; Sequence 1, Application US/08464514
; Patent No. 6265173
; GENERAL INFORMATION:
; APPLICANT: EVANS, RONALD M.
; APPLICANT: MCKEOWN, MICHAEL B.
; APPLICANT: ORO, ANTHONY E.
; APPLICANT: SEGRAVES, WILLIAM A.
; APPLICANT: YAO, TSO-PANG
; TITLE OF INVENTION: MULTIMERIC FORMS OF MEMBERS OF THE
; TITLE OF INVENTION: STEROID/THYROID SUPERFAMILY OF RECEPTORS WITH THE
; TITLE OF INVENTION: ULTRASPIRACLE RECEPTOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESS: PRETTY, SCHROEDER, BRUEGEMANN & CLARK
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: United States
; ZIP: 90071

```



COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/464,514  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/907,908  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31192  
REFERENCE/DOCKET NUMBER: P41 9321  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 546-4737  
TELEFAX: (619) 546-9392  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2304 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 163..1701  
US-08-464-514-1

Query Match 1.8%; Score 33; DB 4; Length 2304;  
Best Local Similarity 63.0%; Pred. No. 21;  
Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 646 TATTAAGGTTGCTTGACATTTTGAATTTCTTCATGATTTTACGGCTCT 705  
DB 2009 TAGTTTACTTTTACTTTACTTCTGTCGTGACATTTTCGTCGTTTGCTCT 1950

QY 706 TTTCAGCTAAATTTCTT 726  
DB 1949 TTTCAGTTTGATTTTGT 1929

RESULT 41  
US-08-486-403-1/c  
Sequence 1, Application US/08486403  
Patent No. 6281330  
GENERAL INFORMATION:  
APPLICANT: EVANS, RONALD M.  
APPLICANT: MCKEOWN, MICHAEL B.  
APPLICANT: ORO, ANTHONY E.  
APPLICANT: SEGRAVES, WILLIAM A.  
APPLICANT: YAO, TSO-PANG  
TITLE OF INVENTION: MULTIMERIC FORMS OF MEMBERS OF THE  
TITLE OF INVENTION: STEROID/THYROID SUPERFAMILY OF RECEPTORS WITH THE  
TITLE OF INVENTION: ULTRASPIRACLE RECEPTOR  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles  
STATE: California  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,403  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/907,908  
FILING DATE: 02-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31192  
REFERENCE/DOCKET NUMBER: P41 9321  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 546-4737  
TELEFAX: (619) 546-9392  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2304 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 163..1701  
US-08-486-403-1

Query Match 1.8%; Score 33; DB 4; Length 2304;  
Best Local Similarity 63.0%; Pred. No. 21;  
Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 646 TATTAAGGTTGCTTGACATTTTGAATTTCTTCATGATTTTACGGCTCT 705  
DB 2009 TAGTTTACTTTTACTTTACTTCTGTCGTGACATTTTCGTCGTTTGCTCT 1950

QY 706 TTTCAGCTAAATTTCTT 726  
DB 1949 TTTCAGTTTGATTTTGT 1929

RESULT 42  
US-09-149-476-221/c  
Sequence 221, Application US/09149476  
Patent No. 6420526  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 186 Human Secreted proteins  
FILE REFERENCE: P2002P1  
CURRENT APPLICATION NUMBER: US/09/149,476  
CURRENT FILING DATE: 1998-09-08  
EARLIER APPLICATION NUMBER: PCT/US98/04493  
EARLIER FILING DATE: 1998-03-06  
EARLIER APPLICATION NUMBER: 60/040,162  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,333  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/038,621  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,626  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,334  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,336  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,163  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/047,600  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,615  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,597  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,502  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,633  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,583



EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057,650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049,610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061,060  
EARLIER FILING DATE: 1997-10-02

Query Match 1.7%; Score 32.6; DB 4; Length 504;  
Best Local Similarity 51.4%; Pred. No. 14;  
Matches 71; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1571 AGCGGACGGTAATGACAGACGTAACCTATACAGTTTGTAAATGCTTTGTTATA 1630  
DB 161 ATCGTAGATTAATGAATGAATGAATCTGAAAGTTTAACTTTCTTTT 102  
QY 1631 GGGGCTCTCTGCTTTAATTTGATTTATATGACATAGGATCTTCTAGAACT 1690  
DB 101 TAGTTTTTCTTTTGTGATTTTTTTTTTTTGGGGTTTGAGTGTGTTGAGGT 42  
QY 1691 TTATGCTCTTTATATT 1708  
DB 41 TTTTNNCCCTTTTTT 24

RESULT 43  
US-09-008-979A-3  
Sequence 3, Application US/09008979A  
Patent No. 6080914

GENERAL INFORMATION:  
APPLICANT: Conner, Timothy W.  
TITLE OF INVENTION: Strawberry Promoters and Genes  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Carmen Rodriguez, Paralegal, Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/008,979A  
FILING DATE:

CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Kammerer, Patricia A.  
REGISTRATION NUMBER: 29,775  
REFERENCE/DOCKET NUMBER: MOBT:058 (38-21(10613)A)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (713) 787-1438  
TELEFAX: (713) 789-2679  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1260 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double

TOPOLOGY: linear  
US-09-008-979A-3

Query Match 1.7%; Score 32.4; DB 3; Length 1260;  
Best Local Similarity 50.6%; Pred. No. 24;  
Matches 78; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1390 AGCATTCTTTTGAAGAAATTCAAATGATTTGCAAGCATTGCTTTAATGAC 1449  
DB 1069 ACCGATTGGAAGCGAGATATTCAGTCGATCGACCAAGTTCCTCCCTCATTTGA 1128  
QY 1450 TGTATGGGGTATTCATTAAACAATCTTTGATCTTTGATTCGTAAGATG 1509  
DB 1129 TGCCGCTGGATCCCAATGAACTGATTTTTTTTAAATTTTCTTAAAGAACCCCG 1188  
QY 1510 CCAAGGAACCTTTTACCCCTATGATGTTTGA 1543  
DB 1189 AGACTGGCTTGCTTGCTTATGATCTTCA 1222

RESULT 44  
US-09-460-618-3  
Sequence 3, Application US/09460618  
Patent No. 6235482

GENERAL INFORMATION:  
APPLICANT: Conner, Timothy W.  
TITLE OF INVENTION: Strawberry Promoters and Genes  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Carmen Rodriguez, Paralegal, Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/460,618  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/008,979  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Kammerer, Patricia A.  
REGISTRATION NUMBER: 29,775  
REFERENCE/DOCKET NUMBER: MOBT:058 (38-21(10613)A)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (713) 787-1438  
TELEFAX: (713) 789-2679  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1260 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-460-618-3

Query Match 1.7%; Score 32.4; DB 4; Length 1260;  
Best Local Similarity 50.6%; Pred. No. 24;  
Matches 78; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1390 AGCATTCTTTTGAAGAAATTCAAATGATTTGCAAGCATTGCTTTAATGAC 1449  
DB 1069 ACCGATTGGAAGCGAGATATTCAGTCGATCGACCAAGTTCCTCCCTCATTTGA 1128  
QY 1450 TGTATGGGGTATTCATTAAACAATCTTTGATCTTTGATTCGTAAGATG 1509  
DB 1129 TGCCGCTGGATCCCAATGAACTGATTTTTTTTAAATTTTCTTAAAGAACCCCG 1188

Qy 1510 CCAAGCGAACCTGTTTACCCCTATGCATGTTTAA 1543  
Db 1189 AGACTGGCCTCTGCTTTGCTCTATGTATCCTTCA 1222

## RESULT 45

US-09-310-235B-3  
; Sequence 3, Application us/09310235B  
; Patent No. 6392030  
; GENERAL INFORMATION:  
; APPLICANT: Conner, Timothy  
; TITLE OF INVENTION: STRAWBERRY PROMOTERS AND GENES  
; FILE REFERENCE: MOBT058--1 (11899.0058.DVUS00)  
; CURRENT APPLICATION NUMBER: US/09/310.235B  
; CURRENT FILING DATE: 1998-05-12  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3  
; LENGTH: 1260  
; TYPE: DNA  
; ORGANISM: fragaria x ananassa  
US-09-310-235B-3

Query Match 1.7%; Score 32.4; DB 4; Length 1260;  
Best Local Similarity 50.6%; Pred. No. 24;  
Matches 78; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 1390 AGCATTTCCTTTTGAAGAAAATTCAAATAGATTGCAAGCCATTGGTGTCTTAAATGAC 1449  
Db 1069 ACCGATTGGAAGCGAGCATATCCAGTCGCATCAGCCAGTGTTCTCCCTCATTTGGA 1128  
Qy 1450 TGTATTGGGGTATTCAATTAACAATACTTTGATCATTTTGTATCGTATTCGTGAAGATCG 1509  
Db 1129 TGCGTCTGGCATCCAATGAACGTGATTTTAAATTTTCTTATAAAGAACCCCG 1188  
Qy 1510 CCAAGCGAACCTGTTTACCCCTATGCATGTTTAA 1543  
Db 1189 AGACTGGCCTCTGCTTTGCTCTATGTATCCTTCA 1222

Search completed: December 17, 2002, 08:13:21  
Job time : 1000 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 09:05:16 ; Search time 1972 Seconds  
(without alignments)  
15308.513 Million cell updates/sec

Title: US-09-868-987-1  
Perfect score: 1864  
Sequence: 1 atggacttcgcattgtc.....atgacaaattcagataatgc 1864

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Result No.	Score	Query Match	Length	DB ID	Description
c 1	22	1.2	595	14	BQ038573
c 2	22	1.2	614	17	BH616209
c 3	22	1.2	647	9	AJ396413
c 4	21	1.1	242	17	AZ781921
c 5	21	1.1	533	10	BB283826
c 6	20	1.1	171	9	A1637310

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

AI637344 DB# 362 G	171	1.1	20	c 7	7
BF113513 EST441103	302	1.1	20	8	
AI447978 mr16d03.x	434	1.1	20	9	
AQ338598 HS_2202.B	449	1.1	20	10	
BI210777 EST528817	497	1.1	20	11	
AL744698 Danio rer	507	1.1	20	12	
BH017124 TDGCT72TH	558	1.1	20	13	
AL52287 Anopheles	690	1.1	20	14	
AL144935 Anopheles	830	1.1	20	15	
BF303297 602030211	893	1.1	20	16	
BF876881 PM4-ET015	918	1.1	19	17	
AL312324 Tetraodon	169	1.0	19	18	
AV223805 AV223805	205	1.0	19	19	
BH342232 CH230-9J1	236	1.0	19	20	
BE527201 M67F07STM	246	1.0	19	21	
AI445399 t124h08.x	262	1.0	19	22	
AV155678 AV155678	286	1.0	19	23	
AV138833 AV138833	298	1.0	19	24	
AI55340 UI-R-C2P-	299	1.0	19	25	
AV146527 AV146527	306	1.0	19	26	
AV117891 AV117891	319	1.0	19	27	
AA081640 zn20C03.f	332	1.0	19	28	
AI466079 vw39d07.y	350	1.0	19	29	
AW436624 77002.MAR	353	1.0	19	30	
BF364126 QVO-NN102	356	1.0	19	31	
AZ537688 AST-DHW(7	367	1.0	19	32	
BQ528747 3524.1_37	376	1.0	19	33	
BJ541842 BJ541842	389	1.0	19	34	
T27099 NBT285H12R	396	1.0	19	35	
BJ529530 BJ529530	411	1.0	19	36	
AA822992 vw39d07.f	419	1.0	19	37	
AZ290933 RFC1-23-1	430	1.0	19	38	
AQ359339 HS_5030.B	430	1.0	19	39	
BH864257 SALK_0956	447	1.0	19	40	
AV968816 AV968816	455	1.0	19	41	
AA798200 ub60e03.f	459	1.0	19	42	
AQ698248 HS_5550.B	459	1.0	19	43	
AI159020 vz86f10.f	461	1.0	19	44	
AZ463779 IM027A19	481	1.0	19	45	
BH177544 010_E_03-	484	1.0	19	46	
W79626 zd81g08.r1	484	1.0	19	47	
W90044 zh69a11.r1	505	1.0	19	48	
BI464770 603202218	506	1.0	19	49	
R55686 y988e12.r1	506	1.0	19	50	
BM935060 UI-M-BH3-	508	1.0	19	51	
AA986773 uc74b05.x	511	1.0	19	52	
BM194502 TCARP2Q11	518	1.0	19	53	
AW012996 SSELFO3.W	520	1.0	19	54	
BF727747 1000052C0	521	1.0	19	55	
BM795924 K-EST0078	524	1.0	19	56	
BQ296942 88036C03	525	1.0	19	57	
BQ297083 88038A02	526	1.0	19	58	
AL703412 DKFZp886C	531	1.0	19	59	
AW732327 B502H04.Y	531	1.0	19	60	
AZ430264 IM0214K08	531	1.0	19	61	
BQ287551 i166e09.y	538	1.0	19	62	
AL695384 DKFZp313N	539	1.0	19	63	
AA518311 vi06C05.f	540	1.0	19	64	
BG338104 602435939	543	1.0	19	65	
BQ287534 i166d04.y	550	1.0	19	66	
AV058003 AV058003	552	1.0	19	67	
BF330723 RC4-BT031	552	1.0	19	68	
BG423960 602447417	553	1.0	19	69	
BH114888 RRC1-24-3	561	1.0	19	70	
AW988655 ugi0h12.y	567	1.0	19	71	
AI425414 my16e05.y	571	1.0	19	72	
AY106979 Zea mayB	571	1.0	19	73	
BH762059 BMBAC326G	586	1.0	19	74	
BG227572 Kc10H01.y	590	1.0	19	75	
AZ430692 IM0215117	593	1.0	19	76	
BH450071 BOHEO31TF	600	1.0	19	77	
AW249380 2819502.5	602	1.0	19	78	
			19	79	

C 80	19	1.0	606	9	AI993818	701514781	C 153	18	1.0	169	9	AUI85971	AUI85971
81	19	1.0	606	9	AA222578	my16e05.r	154	18	1.0	192	10	BE051996	BE051996
82	19	1.0	610	12	BF518890	EST456422	155	18	1.0	198	9	AV163828	AV163828
C 83	19	1.0	611	17	AZ61868	IM0267J21	156	18	1.0	207	17	AZ575764	AZ575764
84	19	1.0	624	17	AQ038180	CITBI-EI-	157	18	1.0	220	13	BI121623	BI121623
85	19	1.0	632	17	AQ379336	RPC11-16	158	18	1.0	223	12	BE856208	BE856208
C 86	19	1.0	639	17	AG141283	Pan trogl	159	18	1.0	225	9	AUI257097	AUI257097
87	19	1.0	649	17	AZ573164	314PvD02	160	18	1.0	228	9	AI827328	AI827328
C 88	19	1.0	653	17	AZ259849	RPC1-23-1	161	18	1.0	229	12	BG013808	BG013808
C 89	19	1.0	665	17	AQ480736	RPC1-11-2	162	18	1.0	237	9	AV218287	AV218287
C 90	19	1.0	667	17	BE536878	601064885	163	18	1.0	246	9	AA366819	AA366819
C 91	19	1.0	677	12	BF714797	mab03d03-	164	18	1.0	248	12	BE986213	BE986213
92	19	1.0	678	10	BE264998	601193827	165	18	1.0	253	10	BE068699	BE068699
93	19	1.0	685	9	AL503713	AL503713	166	18	1.0	264	12	BE984790	BE984790
94	19	1.0	686	14	BQ597373	MI-P-A2-a	167	18	1.0	267	9	AV094338	AV094338
95	19	1.0	692	12	BF971071	602270624	168	18	1.0	269	12	BF326607	BF326607
C 96	19	1.0	694	17	AZ199764	SP 1040 A	169	18	1.0	271	9	AV092510	AV092510
C 97	19	1.0	695	17	AZ524062	226PBG07	170	18	1.0	271	9	AV141665	AV141665
C 98	19	1.0	713	17	BH292163	CH230-186	171	18	1.0	272	10	AW933803	AW933803
C 99	19	1.0	714	17	AG009030	HOMO_BAP1	172	18	1.0	275	9	AJ500448	AJ500448
100	19	1.0	728	12	BG697501	602660695	173	18	1.0	276	9	AJ500137	AJ500137
101	19	1.0	731	13	BI755220	60303227	174	18	1.0	281	9	AA484891	AA484891
102	19	1.0	733	17	BH181011	017 H 06-	175	18	1.0	281	10	BB022698	BB022698
103	19	1.0	743	17	CNS07MCP	AL617961 T7 end of	176	18	1.0	281	14	BO696513	BO696513
104	19	1.0	743	17	BH523054	HOG167TR	177	18	1.0	289	17	AZ252904	AZ252904
105	19	1.0	743	12	BG715477	602675447	178	18	1.0	293	9	AA12301	AA12301
C 106	19	1.0	751	9	AI125363	AI125363	179	18	1.0	294	14	Z14535	Z14535
107	19	1.0	764	13	BI758569	603023228	180	18	1.0	297	10	AV817576	AV817576
C 108	19	1.0	771	13	BI753921	603022928	181	18	1.0	299	9	AA921709	AA921709
C 109	19	1.0	780	9	AUI17917	AUI17917	182	18	1.0	310	14	M56262	M56262
110	19	1.0	796	12	BG699238	602678909	183	18	1.0	311	9	AI014820	AI014820
111	19	1.0	796	12	BG699445	602679138	184	18	1.0	316	9	AI702276	AI702276
C 112	19	1.0	806	9	AUI11827	AUI11827	185	18	1.0	319	10	AV932928	AV932928
113	19	1.0	807	13	BI648132	603278450	186	18	1.0	320	9	AI511305	AI511305
114	19	1.0	812	9	AUI43540	AUI43540	187	18	1.0	321	9	AV009727	AV009727
C 115	19	1.0	815	9	AUI43540	AI126380	188	18	1.0	329	12	BE221229	BE221229
116	19	1.0	827	9	AUI26380	AUI26380	189	18	1.0	330	14	D74236	D74236
117	19	1.0	833	12	BG704268	602687643	190	18	1.0	331	10	AW931706	AW931706
118	19	1.0	842	12	BG697733	602661209	191	18	1.0	332	10	BB315355	BB315355
119	19	1.0	845	12	BG870872	602792283	192	18	1.0	333	10	BE646309	BE646309
120	19	1.0	862	12	BG714892	602675746	193	18	1.0	333	12	BF447278	BF447278
C 121	19	1.0	872	9	AUI32302	AUI32302	194	18	1.0	340	10	AW151093	AW151093
122	19	1.0	873	12	BE912645	601663688	195	18	1.0	340	10	AW279831	AW279831
123	19	1.0	881	12	BE788310	601480113	196	18	1.0	341	9	AUI10627	AUI10627
124	19	1.0	883	14	BQ961422	AGENCOURT	197	18	1.0	342	12	BG225557	BG225557
125	19	1.0	886	12	BG765535	602739450	198	18	1.0	344	14	T69881	T69881
126	19	1.0	890	14	BQ226575	AGENCOURT	199	18	1.0	350	9	AUI11156	AUI11156
127	19	1.0	903	13	BI560672	603254701	200	18	1.0	350	9	AUI12269	AUI12269
128	19	1.0	905	17	AZ543188	EMTGR34TR	201	18	1.0	352	9	AA930406	AA930406
129	19	1.0	912	12	BG031637	602299785	202	18	1.0	360	9	AJ467982	AJ467982
C 130	19	1.0	916	12	BF103352	BI462668	203	18	1.0	360	9	AV168982	AV168982
131	19	1.0	951	13	BI462668	603202109	204	18	1.0	360	9	AV189465	AV189465
132	19	1.0	957	12	BG036197	602326945	205	18	1.0	360	14	C09172	C09172
133	19	1.0	969	9	AL514362	AL514362	206	18	1.0	360	14	C40460	C40460
134	19	1.0	978	9	AL545110	AL545110	207	18	1.0	360	14	C40830	C40830
135	19	1.0	984	13	BI687550	603312562	208	18	1.0	360	14	C42115	C42115
136	19	1.0	1062	10	BE617480	601442138	209	18	1.0	360	14	C43462	C43462
137	19	1.0	1062	12	CNS03908	Te1triodon	210	18	1.0	360	14	D67729	D67729
138	19	1.0	1075	12	BG035417	602325044	211	18	1.0	360	14	D69369	D69369
139	19	1.0	1086	14	BM924516	AGENCOURT	212	18	1.0	360	14	D70639	D70639
140	19	1.0	1088	14	BM477275	AGENCOURT	213	18	1.0	360	14	D74316	D74316
141	19	1.0	1098	14	BQ421029	AGENCOURT	214	18	1.0	360	14	D74393	D74393
142	19	1.0	1114	13	BM462885	AGENCOURT	215	18	1.0	360	14	D74404	D74404
143	19	1.0	1142	13	BM547681	AGENCOURT	216	18	1.0	360	14	D74759	D74759
144	19	1.0	1144	13	BM547681	AGENCOURT	217	18	1.0	360	14	D75017	D75017
145	19	1.0	1692	12	BG025559	602274525	218	18	1.0	360	14	D75884	D75884
C 146	19	1.0	1890	13	BM416552	CE21649 M	219	18	1.0	360	14	D75957	D75957
147	19	1.0	2646	11	BC004332	Homo_sapi	220	18	1.0	361	9	AI974743	AI974743
C 148	19	1.0	42	17	TAZ55A05P	TAZ55A05P	221	18	1.0	365	17	AZ886610	AZ886610
C 149	19	1.0	92	17	AL766294	Arabidops	222	18	1.0	366	10	BE496520	BE496520
150	19	1.0	129	17	BH454374	BE06S25TR	223	18	1.0	371	10	BE209089	BE209089
C 151	19	1.0	138	17	BF641717	NF066C08I	224	18	1.0	374	12	BF115459	BF115459
C 152	19	1.0	143	17	AO471195	CITBI-EI-	225	18	1.0	375	10	BE008957	BE008957
			168	13	BI189418	e4ff11fs.f							







518	18	1.0	670	13	BJ092159	591	18	1.0	764	14	BM816201	BM816201	HC01C09.T
519	18	1.0	672	14	BQ004212	c 592	18	1.0	767	17	AQ855319	AQ855319	CGG1851B
520	18	1.0	672	14	BH004565	c 593	18	1.0	776	13	BI931308	BI931308	EST551197
521	18	1.0	672	17	AG178668	c 594	18	1.0	778	12	BG584580	BG584580	EST486341
522	18	1.0	673	10	AV928565	595	18	1.0	778	12	BF160869	BF160869	601770880
523	18	1.0	673	14	BQ660126	c 596	18	1.0	779	17	BH086307	BH086307	RPCI-24-2
524	18	1.0	674	9	AJ781444	c 597	18	1.0	780	13	BM412340	BM412340	EST586667
525	18	1.0	675	13	BI928689	c 598	18	1.0	784	9	AJ451545	AJ451545	AJ451545
526	18	1.0	677	10	AV922390	c 599	18	1.0	787	17	BH448401	BH448401	BOGGR79TF
527	18	1.0	677	13	BM249992	600	18	1.0	788	14	BQ879008	BQ879008	AGENCOURT
528	18	1.0	677	12	BF465556	601	18	1.0	792	17	BH521792	BH521792	BOGPF12TF
529	18	1.0	679	12	AV943971	c 602	18	1.0	792	17	BH734833	BH734833	BOMPF76TF
530	18	1.0	680	12	EG416629	c 603	18	1.0	793	12	BG206299	BG206299	RST25744
531	18	1.0	681	12	EG416629	c 604	18	1.0	793	17	BH681801	BH681801	BOHUF88TR
532	18	1.0	683	12	BG073020	c 605	18	1.0	794	9	AJ446286	AJ446286	AJ446286
533	18	1.0	684	13	BI954326	606	18	1.0	794	12	BI913092	BI913092	601179695
534	18	1.0	686	10	AV931795	607	18	1.0	796	13	BG786167	BG786167	SEAUMC006
535	18	1.0	686	13	BI931857	c 608	18	1.0	796	17	CNS048NZ	CNS048NZ	Tetrarodon
536	18	1.0	687	10	AV920880	c 609	18	1.0	797	10	BE282966	BE282966	601100467
537	18	1.0	688	17	AG120635	c 610	18	1.0	800	17	AQ741766	AQ741766	HS 5566_B
538	18	1.0	689	13	BI274290	c 611	18	1.0	801	17	BH556535	BH556535	BOGJ153TF
539	18	1.0	691	10	AW647649	612	18	1.0	812	14	BQ014512	BQ014512	UT-H-ED1-
540	18	1.0	692	10	BE061053	c 613	18	1.0	812	17	AQ781379	AQ781379	HS 3117_A
541	18	1.0	692	13	BM426428	614	18	1.0	815	14	BM816125	BM816125	HC107E04
542	18	1.0	693	13	BJ173851	c 615	18	1.0	816	17	AJ448226	AJ448226	AV922164
543	18	1.0	694	17	AG053115	c 616	18	1.0	821	10	BE622399	BE622399	601441189
544	18	1.0	696	9	AI822624	617	18	1.0	821	12	BF257008	BF257008	HVSM8F001
545	18	1.0	697	13	BI282003	618	18	1.0	823	12	BF257053	BF257053	HVSM8F001
546	18	1.0	697	10	AV922164	c 619	18	1.0	824	9	AJ448226	AJ448226	AV922164
547	18	1.0	697	10	AW775266	c 620	18	1.0	827	17	BH506273	BH506273	BOGUP95TR
548	18	1.0	697	14	QJ198432	621	18	1.0	830	12	BF627418	BF627418	HVSM8B000
549	18	1.0	698	10	AV928199	c 622	18	1.0	831	12	BG309392	BG309392	HVSM8C000
550	18	1.0	699	10	AV927418	c 623	18	1.0	831	13	BI756513	BI756513	603024652
551	18	1.0	699	14	BQ765162	624	18	1.0	832	10	BE412827	BE412827	MCG008_B1
552	18	1.0	700	9	AL503382	c 625	18	1.0	833	17	BH156840	BH156840	ENTQY92TF
553	18	1.0	700	9	AL503468	626	18	1.0	834	12	BF5311753	BF5311753	602072653
554	18	1.0	700	9	AL503633	627	18	1.0	835	12	BG583345	BG583345	EST484895
555	18	1.0	700	9	AL506005	628	18	1.0	835	17	AZ539178	AZ539178	ENTDU18TR
556	18	1.0	701	12	BF027488	629	18	1.0	840	9	AU119509	AU119509	AU119509
557	18	1.0	701	14	BQ601991	630	18	1.0	840	12	BF255569	BF255569	HVSM8F000
558	18	1.0	702	10	AV935071	631	18	1.0	841	12	BG343018	BG343018	HVSM8G000
559	18	1.0	702	17	BH167966	632	18	1.0	843	17	AZ527901	AZ527901	ENTCM63TF
560	18	1.0	703	10	AV930346	c 633	18	1.0	844	12	BF256474	BF256474	HVSM8F001
561	18	1.0	704	10	AV921469	634	18	1.0	848	12	BF136816	BF136816	601782294
562	18	1.0	704	12	BG128722	635	18	1.0	862	12	BF264412	BF264412	HV C8a000
563	18	1.0	705	13	BI958749	c 636	18	1.0	866	9	AL665647	AL665647	AL665647
564	18	1.0	705	13	BM437647	637	18	1.0	868	13	BI524800	BI524800	603051794
565	18	1.0	706	17	AZ336819	c 638	18	1.0	869	17	CNS05MMH	CNS05MMH	Tetrarodon
566	18	1.0	707	14	BQ767507	639	18	1.0	870	12	BF253769	BF253769	HVSM8F000
567	18	1.0	708	9	AU078033	640	18	1.0	870	17	BH163366	BH163366	ENTTQ37TR
568	18	1.0	709	12	BG593890	641	18	1.0	871	12	BG758855	BG758855	602713246
569	18	1.0	709	12	BG838243	642	18	1.0	872	12	BG344675	BG344675	HVSM8G001
570	18	1.0	709	13	BJ083312	c 643	18	1.0	872	13	BI260399	BI260399	602969465
571	18	1.0	710	10	AV846750	644	18	1.0	873	9	AU117264	AU117264	AU117264
572	18	1.0	712	12	BG125341	c 645	18	1.0	877	10	BE412409	BE412409	MCG004_H0
573	18	1.0	715	9	AJ399445	646	18	1.0	877	12	BF261703	BF261703	HV C8a000
574	18	1.0	716	17	AQ925439	647	18	1.0	877	17	AZ668313	AZ668313	ENTMT56TF
575	18	1.0	716	17	AZ239697	c 648	18	1.0	878	17	AZ672110	AZ672110	ENTML25TR
576	18	1.0	717	13	BI178338	649	18	1.0	879	9	AL668616	AL668616	AL668616
577	18	1.0	718	13	BI928227	c 650	18	1.0	881	17	AZ693229	AZ693229	ENTTF60TF
578	18	1.0	719	10	AV704003	651	18	1.0	886	12	BG344983	BG344983	HVSM8G001
579	18	1.0	719	10	AV704298	652	18	1.0	888	12	BF253735	BF253735	HVSM8F000
580	18	1.0	719	17	AZ612882	653	18	1.0	889	14	BQ941423	BQ941423	AGENCOURT
581	18	1.0	722	12	BG126718	c 654	18	1.0	889	17	BH131529	BH131529	ENTNFS2TR
582	18	1.0	728	17	AQ357201	655	18	1.0	892	17	AQ900774	AQ900774	HS 3179_B
583	18	1.0	732	9	AJ451084	656	18	1.0	892	17	AZ543030	AZ543030	ENTD062TR
584	18	1.0	739	17	BH517501	657	18	1.0	894	14	BQ440956	BQ440956	AGENCOURT
585	18	1.0	746	13	BI927089	658	18	1.0	895	12	BF309892	BF309892	601892358
586	18	1.0	751	17	BH045483	c 659	18	1.0	895	17	AZ684429	AZ684429	ENTJN19TF
587	18	1.0	758	9	AJ451391	c 660	18	1.0	896	17	AZ540035	AZ540035	ENTDGB9TR
588	18	1.0	758	12	BG838711	c 661	18	1.0	897	17	AZ681234	AZ681234	ENTKGI6TF
589	18	1.0	758	17	BI122568	662	18	1.0	898	17	BH138611	BH138611	ENTND07TF
590	18	1.0	763	12	BG526124	663	18	1.0	899	17	BH158182	BH158182	ENTTR10TF





```

956 17 0.9 341 10 BE575543 dcs8d02.x
957 17 0.9 342 10 AV954787 AV954787
958 17 0.9 342 10 AM289438 AM289438
959 17 0.9 343 10 A1283127 A1283127
960 17 0.9 343 10 AM093948 AM093948
961 17 0.9 343 10 AA687187 AA687187
962 17 0.9 345 10 A1143657 A1143657
963 17 0.9 345 10 T07776 T07776
964 17 0.9 346 13 BM168523 BM168523
965 17 0.9 347 14 C90429 C90429
966 17 0.9 349 9 AA934410 AA934410
967 17 0.9 349 9 A1212731 A1212731
968 17 0.9 349 9 A1944550 A1944550
969 17 0.9 350 10 A1158187 A1158187
970 17 0.9 350 10 AM197747 AM197747
971 17 0.9 350 10 BM65794 BM65794
972 17 0.9 352 17 AZ070422 AZ070422
973 17 0.9 353 10 AM706556 AM706556
974 17 0.9 355 10 AM439949 AM439949
975 17 0.9 355 14 HA8561 HA8561
976 17 0.9 359 10 AM886684 AM886684
977 17 0.9 359 12 BF221984 BF221984
978 17 0.9 359 13 BM184924 BM184924
979 17 0.9 360 9 AA813897 AA813897
980 17 0.9 360 14 R15202 R15202
981 17 0.9 360 17 AZ891026 AZ891026
982 17 0.9 361 9 AA690062 AA690062
983 17 0.9 361 13 B1496085 B1496085
984 17 0.9 362 9 AA981996 AA981996
985 17 0.9 362 9 AL804921 AL804921
986 17 0.9 362 9 AA425420 AA425420
987 17 0.9 362 10 AA451518 AA451518
988 17 0.9 362 12 BF115345 BF115345
989 17 0.9 362 14 BQ844416 BQ844416
990 17 0.9 363 10 A0288471 A0288471
991 17 0.9 363 17 A0634984 A0634984
992 17 0.9 364 9 A1372036 A1372036
993 17 0.9 364 12 BE500966 BE500966
994 17 0.9 365 12 BG000124 BG000124
995 17 0.9 365 14 H78521 H78521
996 17 0.9 366 9 A1004695 A1004695
997 17 0.9 367 9 AA492434 AA492434
998 17 0.9 367 10 AV980531 AV980531
999 17 0.9 367 13 B1924752 B1924752
1000 17 0.9 368 9 A1270597 A1270597

```

## ALIGNMENTS

```

RESULT 1
BQ038573/c 595 bp mRNA linear EST 01-MAY-2002
LOCUS BQ038573
DEFINITION BQ038573
            BQ038573.2 GI:20383326
            EST.
            chicken.
            Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
            1 (bases 1 to 595)
            Morgan, R. W. and Burnside, J.
            Chicken lymphoid ESTs
            Unpublished (2001)
            On Mar 27, 2002 this sequence version replaced gi:19772113.
            Contact: Joan Burnside
            Molecular Endocrinology
            University of Delaware

```

```

FEATURES
    source
        1..595
        /organism="Gallus gallus"
        /db_xref="taxon:9031"
        /clone_lib="normalized chicken lymphoid cDNA library"
        /tissue_type="thymus, bursa, spleen, PBL, bone marrow"
        /lab_host="E.coli EMDH108"
        /note="Vector: PCMVSPORT 6"

```

```

BASE COUNT 128 a 168 c 161 g 120 t 18 others
ORIGIN

```

```

Query Match 1.2% Score 22; DB 14; Length 595;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 22; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1277 GCTTGGCAATCATCTGCTT 1298
Db 462 GCTTGGCAATCATCTGCTT 441

```

```

RESULT 2
BHE16209 614 bp DNA linear GSS 28-JAN-2002
LOCUS BHE16209
DEFINITION BHE16209
            BHE16209.1 GI:18380897
            GSS.
            Brugia malayi.
            Brugia malayi.
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Brugia.
            1 (bases 1 to 614)
            Whitton, C., Daub, J., Ware, J., Quail, M., Hall, N., Barrell, B., Foster
            J., Guillian, D., Slatko, B. and Blaxter, M.
            Genome survey sequences from the human parasitic nematode Brugia
            malayi.
            Unpublished (2000)
            Contact: Blaxter ML
            Institute of Cell, Animal and Population Biology
            University of Edinburgh
            Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
            3JT, UK
            Tel: +44 131 650 6760
            Fax: +44 131 670 5450
            Email: mark.blaxter@ed.ac.uk

```

```

REFERENCE
    Whitton, C., Daub, J., Ware, J., Quail, M., Hall, N., Barrell, B., Foster
    J., Guillian, D., Slatko, B. and Blaxter, M.
    Genome survey sequences from the human parasitic nematode Brugia
    malayi.
    Unpublished (2000)
    Contact: Blaxter ML
    Institute of Cell, Animal and Population Biology
    University of Edinburgh
    Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
    3JT, UK
    Tel: +44 131 650 6760
    Fax: +44 131 670 5450
    Email: mark.blaxter@ed.ac.uk

```

```

FEATURES
    source
        1..614
        Location/Qualifiers
        1..614
        /organism="Brugia malayi"
        /strain="TBS"
        /db_xref="taxon:6279"
        /clone_lib="Brugia malayi Genomic Bac Library 3"
        /sex="Mixed (male and female)"
        /tissue_type="whole parasite"
        /dev_stage="microfilaria (L1)"
        /note="Vector: pBAC3.6; Site 1: BamH I; Brugia malayi
        genomic DNA was partially cleaved with Sau3A I and size
        fractionated. 7,392 clones were generated with mean insert
        size ~48 kbp. The library was constructed by Claire
        Whitton, Blaxter Nematode Genetics Lab, University of

```

Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing Unit, the Sanger Centre, Cambridge, UK."

BASE COUNT 187 a 108 c 103 g 216 t  
 ORIGIN  
 Query Match 1.2%; Score 22; DB 17; Length 614;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1392 CACATTTCTTTTGAAGAAAT 1413  
 |||||  
 Db 217 CACATTTCTTTTGAAGAAAT 238

RESULT 3  
 AJ396413/c  
 LOCUS  
 DEFINITION AJ396413 dkfz426 Gallus gallus cDNA clone 26f1r1, mRNA sequence.  
 ACCESSION AJ396413  
 VERSION AJ396413.1 GI:7128327  
 KEYWORDS EST.  
 SOURCE chicken.

ORGANISM Gallus gallus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 1 (bases 1 to 647)

REFERENCE  
 AUTHORS Abdrakhmanov, I., Lodygin, D., Gerth, P., Arakawa, H., Law, A., Plachy, J., Korn, B. and Buerstedde, J. M.  
 TITLE A large database of chicken bursal ESTs as a resource for the analysis of vertebrate gene function  
 JOURNAL Genome Res. 10 (12), 2062-2069 (2000)  
 MEDLINE 20568495  
 COMMENT Contact: Buerstedde JM  
 Cellular Immunology  
 Heinrich-Pette-Institut  
 Martinistr. 52, 20351 Hamburg, Germany  
 Email: URL: <http://Genetics.hpi.uni-hamburg.de/dt40est.html>.

FEATURES  
 source  
 1. 647  
 /organism="Gallus gallus"  
 /strain="CB"  
 /db\_xref="taxon:9031"  
 /clone="26f2r1"  
 /clone\_lib="dkfz426"  
 /tissue\_type="Bursa of Fabricius"  
 /cell\_type="Bursal lymphocyte"  
 /dev\_stage="2-3 weeks old"  
 BASE COUNT 137 a 187 c 191 g 132 t  
 ORIGIN

Query Match 1.2%; Score 22; DB 9; Length 647;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1277 GCTTTGGCAATCATCTTCTCT 1298  
 |||||  
 Db 435 GCTTTGGCAATCATCTTCTCT 414

RESULT 4  
 AZ781921  
 LOCUS  
 DEFINITION 2M0021L08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0021L08 R, DNA sequence.  
 ACCESSION AZ781921  
 VERSION AZ781921.1 GI:12915097  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE  
 1 (bases 1 to 242)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

## TITLE

Mouse whole genome scaffolding with paired end reads from 10kb

## JOURNAL

Unpublished (2000)

## COMMENT

Contact: Robert B. Weiss  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA

Tel: 801 585 5606  
 Fax: 801 585 7177

Email: [cdunn@genetics.utah.edu](mailto:cdunn@genetics.utah.edu)

Insert Length: 10000 Std Error: 0.00

Plate: 0021 row: L column: 08

Seq primer: CACACAGGAACACGCTATGACC

Class: plasmid ends

High quality sequence stop: 242.

## FEATURES

## source

1. 242  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC2M0021L08"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Ti-resistant, F-"  
 /notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gl|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 106 a 31 c 33 g 72 t  
 ORIGIN

Query Match 1.1%; Score 21; DB 17; Length 242;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 TGAATAATCATGCCAGTGTCT 152  
 |||||  
 Db 44 TGAATAATCATGCCAGTGTCT 64

RESULT 5  
 BB283826/c  
 LOCUS

DEFINITION BB283826 RIKEN full-length enriched, adult retina Mus musculus cDNA clone A930103H22 3', mRNA sequence.

ACCESSION BB283826  
 VERSION BB283826.1 GI:8984275  
 KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE  
 1 (bases 1 to 533)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.



ACCESSION	A1637344	GI:4689578
VERSION	A1637344.1	
KEYWORDS		
SOURCE	tomato.	
ORGANISM	Lycopersicon esculentum	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum	





Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1473 ATACTTGATCATTTTGAT 1492  
|||||  
Db 414 ATACTTGATCATTTTGAT 433

RESULT 10  
LOCUS A0338598 449 bp DNA linear GSS 12-JUN-1999  
DEFINITION HS 2202 B1 C01 MR CIT Approved Human Genomic Sperm Library D Homo  
Sapiens genomic clone Plate=2202 Col=1 Row=F, DNA sequence.  
ACCESSION A0338598  
VERSION A0338598.1 GI:4144614  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 449)  
Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
CONTACT: Mahairas GG, Wallace JC, Hood L  
99380589  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 2202 row: F column: 1  
Seq primer: M13 Reverse  
Class: BAC ends  
High quality sequence stop: 449.

FEATURES  
source  
1..449  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_plate="2202 Col=1 Row=F"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelBAC11, BAC Clones in  
E-Coli DH10B"

BASE COUNT 133 a 90 c 86 g 139 t 1 others  
ORIGIN

Query Match 1.1%; Score 20; DB 17; Length 449;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1410 AATTCGAATGATTTGCA 1429  
|||||  
Db 171 AATTCGAATGATTTGCA 190

RESULT 11  
LOCUS BI210777 497 bp mRNA linear EST 11-JUN-2001  
DEFINITION EST528817 CTOS Lycopersicon esculentum cDNA clone CTOS24E24 5' end,  
mRNA sequence.  
ACCESSION BI210777  
VERSION BI210777.1 GI:14688501  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterales; eunasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.

REFERENCE 1 (bases 1 to 497)  
AUTHORS van der Hoeven,R., Sun,H., Bezzerides,J., Cho,J., Uterback,R.,  
Romling,C. and Tanksey,S.  
TITLE Generation of ESTs from Tomato Suspension Cultures  
JOURNAL Unpublished (2001)  
CONTACT: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: http://www.genome.clemson.edu/orders/index.html.  
Location/Qualifiers

FEATURES  
source  
1..497  
/organism="Lycopersicon esculentum"  
/cultivar="TA496, E6203"  
/db\_xref="taxon:4081"  
/clone\_plate="CTOS24E24"  
/clone\_lib="CTOS"  
/issue\_type="Suspension cultures"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; Suspension cultures of L.esculentum E6203 were grown  
in Murashige and Skoog based medium, supplemented with 15%  
coconut milk (filter sterilized and added after  
autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).  
Fresh medium was added every 7 days, and cultures were  
grown at 25 C, with 12hrs of light and continuous  
shaking."

BASE COUNT 139 a 84 c 109 g 165 t  
ORIGIN

Query Match 1.1%; Score 20; DB 13; Length 497;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 835 TGGAGGTGTTTCTTTAG 854  
|||||  
Db 371 TGGAGGTGTTTCTTTAG 390

RESULT 12  
LOCUS DR12E9S/c 507 bp DNA linear GSS 06-JUN-2002  
DEFINITION Danio rerio genomic clone DKey-12E9, genomic survey sequence.  
ACCESSION AL744698  
VERSION AL744698.1 GI:21338981  
KEYWORDS GSS.  
SOURCE zebrafish.  
ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 507)

REFERENCE Humphray,S.J., Huckle,E. and Hunt,S.E.  
AUTHORS Direct Submission  
TITLE Submitted (06-JUN-2002) The Sanger Institute, Wellcome Trust Genome  
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:  
humquerry@sanger.ac.uk Unpublished  
COMMENT This sequence was generated from the SP6 end of BAC 12E9. 12E9 is  
part of the Daniokey Pilot BAC library created by R. Plasterk and  
N.V. Keygene.  
Further details: http://www.sanger.ac.uk/Projects/D\_rerio/.

FEATURES  
source  
1..507  
Location/Qualifiers  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone\_plate="DKey-12E9"  
/issue\_type="Testis"  
/note="vector pindigobAC-536"  
BASE COUNT 184 a 76 c 82 g 165 t  
ORIGIN

```

Query Match      1.1%; Score 20; DB 17; Length 507;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1644 TCTTTAAATTTGCATTATT 1663
Db 338 TCTTTAAATTTGCATTATT 319

RESULT 13
LOCUS BH017124 658 bp DNA linear GSS 18-MAY-2001
DEFINITION TDGCT72TH CTG Lycopersicon esculentum genomic clone CTG18L24, DNA
sequence.
ACCESSION BH017124
VERSION BH017124.1 GI:14148162
KEYWORDS GSS,
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 658)
AUTHORS van der Hoeven, R., Sun, H., Cho, J., Utterback, T., Ronning, C. and
Tanksley, S.
TITLE Tomato Demethylated Genomic DNA Sequences
JOURNAL Unpublished (2001)
COMMENT Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
tomato demethylated genomic DNA
Seq primer: M13P-R
Class: shotgun.
FEATURES
source
Location/Qualifiers
1..658
/organism="Lycopersicon esculentum"
/cultivar="E6203"
/db_xref="taxon:4081"
/clone="CTG18L24"
/clone_lib="CTG"
/tissue_type="young leaves"
/dev_stage="12-14 weeks post harvest"
/lab_host="E.coli JM109"
/notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; This library was made from short EcoRI digested
fragments of the genome of Lycopersicon esculentum ligated
into pBS (SK-). The fragments were cloned into the
methylation restrictive E.coli strain JM109 with the
purpose of enriching the library for non-methylated DNA
fragments. This procedure may enrich the pool of cloned
fragments in JM109 cells for sequences representing
expressed genes. Average insert size 1.27 kb."
BASE COUNT 159 a 141 c 143 g 215 t
ORIGIN

Query Match      1.1%; Score 20; DB 17; Length 658;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1199 AACGCCAAATTTTGGTCAA 1218
Db 517 AACGCCAAATTTTGGTCAA 536

RESULT 14
LOCUS CNS01NJ2 690 bp DNA linear GSS 14-JUN-2001
DEFINITION Anopheles gambiae GSS T7 end of clone 24D03 of NotreDamel library
from strain PEST of Anopheles gambiae (African malaria mosquito),

```

```

genomic survey sequence.
AL152287 830 bp DNA linear GSS 06-JUN-2001
VERSION AL152287.1 GI:7013206
KEYWORDS GSS,
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 690)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
REFERENCE 2 (bases 1 to 690)
AUTHORS Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissenbach, J.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
COMMENT This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
FEATURES
source
Location/Qualifiers
1..690
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="24D03"
/clone_lib="NotreDamel"
/notes="end : T7"
BASE COUNT 245 a 105 c 105 g 210 t 25 others
ORIGIN

Query Match      1.1%; Score 20; DB 17; Length 690;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1385 TTTATAGCACATTTCTTTT 1404
Db 658 TTTATAGCACATTTCTTTT 639

RESULT 15
LOCUS CNS01HVA 830 bp DNA linear GSS 06-JUN-2001
DEFINITION Anopheles gambiae GSS SP6 end of clone 09J12 of NotreDamel library
from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION AL144935
VERSION AL144935.1 GI:7003097
KEYWORDS GSS,
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 830)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
REFERENCE 2 (bases 1 to 830)
AUTHORS Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissenbach, J.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
COMMENT This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.

```

```

FEATURES
  source
    Location/Qualifiers
      1..830
        /organism="Mus musculus"
        /strain="FVB/N"
        /db_xref="taxon:10090"
        /clone="IMAGE:4165251"
        /clone_lib="NCI CGAP SG2"
        /lab_host="DH0B (T1 phage-resistant)"
        /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site: 1;
        Note: Site 2: SalI; Cloned unidirectionally. Primer: Oligo
        dT. Average insert size 1.3 kb. Constructed by Life
        Technologies. Note: this is a NCI CGAP Library."
BASE COUNT      372 a      185 c      153 g      182 t      16 others
ORIGIN
Query Match      1.1%; Score 20; DB 12; Length 893;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1400 TTTTGAAGAAATTCAAAT 1419
|||||
Db 653 TTTTGAAGAAATTCAAAT 672

RESULT 17
CNS04XXZF/c

```

```

LOCUS      CNS04XXZF      918 bp      DNA      linear      GSS 26-JUL-2000
DEFINITION Tetradon nigroviridis genome survey, sequence SP6 end of clone
016P16 of library B from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION  AL312324
VERSION    AL312324.1 GI:9545192
KEYWORDS  GSS; genome survey sequence.
SOURCE    Tetradon nigroviridis.
ORGANISM  Tetradon nigroviridis.
REFERENCE 1 (bases 1 to 918)
AUTHORS  Roest Crolius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
ESTIMATE of human gene number provided by genome-wide analysis
using Tetradon nigroviridis DNA sequence
JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE   20296633
PUBMED    10835645
REFERENCE 2 (bases 1 to 918)
AUTHORS  Crolius,H.R., Jalllon,O., Dasilva,C., Orouf-Costaz,C., Fizames,C.,
Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
MEDLINE   20359837
PUBMED    10899143
REFERENCE 3 (bases 1 to 918)
AUTHORS  Genoscope.
Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradon.
Location/Qualifiers
  source
    1..918
      /organism="Tetradon nigroviridis"
      /db_xref="taxon:99883"
      /clone="016P16"
      /clone_lib="B"
      /note="Genoscope sequence ID : COAB01CDH08B1-end : SP6"
BASE COUNT      192 a      208 c      205 g      277 t      36 others
ORIGIN
Query Match      1.1%; Score 20; DB 17; Length 918;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1055 ATCAAAATCTATTAGTGA 1074
|||||
Db 754 ATCAAAATCTATTAGTGA 735

RESULT 18
BF876881/c
LOCUS      BF876881      169 bp      mRNA      linear      EST 17-JAN-2001
DEFINITION PM4-ET0154-151100-004-a12 ET0154 Homo sapiens CDNA, mRNA sequence.
ACCESSION  BF876881
VERSION    BF876881.1 GI:12267011
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 169)
AUTHORS  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Dias Neto,B., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,

```

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&t2=PM4-ET0154-151100-004-a12&t3=2000-11-15&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 169.

## FEATURES

```

source
1. .169
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0154"
/dev_stages="Adult"
/note="Organ: lung_tumor; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORFESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
39 c 43 g 32 t
55 a

```

## BASE COUNT

```

ORIGIN
Query Match 1.0%; Score 19; DB 12; Length 169;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1280 TTGGCAATCATCTGCTCT 1298

Db 47 TTGGCAATCATCTGCTCT 29

## RESULT 19

```

AV223805 205 bp mRNA linear EST 14-NOV-2001
LOCUS AV223805 RIKEN full-length enriched, 18 days pregnant, placenta and extra embryonic tissue Mus musculus cDNA clone 3830414A10 3' similar to AF032668 Rattus norvegicus rsec15 mRNA, mRNA sequence.
AV223805
AV223805.1 GI:6173183

```

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

```

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 205)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominega,N., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki

```

## TITLE

## COMMENT

Unpublished (1999)

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gscc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

## FEATURES

## source

```

1. .205
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="3830414A10"
/clone_lib="RIKEN full-length enriched, 18 days pregnant, placenta and extra embryonic tissue"
/sex="female"
/tissue_type="placenta and extra embryonic tissue"
/dev_stage="18 days pregnant, adult"
/lab_host="DH108"
/note="Site 1: Sall; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 50.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTTAATAATATCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda
```

```

BASE COUNT 77 a 29 c 24 g 75 t
ORIGIN

```

```

Query Match 1.0%; Score 19; DB 9; Length 205;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1400 TTTTGAAGAAAATTCAAA 1418

Db 133 TTTTGAAGAAAATTCAAA 151

## RESULT 20

```

BH342232/c
LOCUS BH342232 236 bp DNA linear GSS 03-DEC-2001
DEFINITION CH230-9J13 TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-9J13, DNA sequence.
ACCESSION BH342232
VERSION BH342232.1 GI:17272966
KEYWORDS GSS.
SOURCE Norway rat.

```

Michigan State University  
 224 Biochemistry, Michigan State University, East Lansing, MI 48824  
 ' USA  
 Tel: 517 355 1609  
 Fax: 517 355 9334  
 Email: demning@msu.edu  
 Clones were originally prepared at Michigan State University.  
 Arabidopsis Biological Resource Center, The Ohio State University,  
 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
 USA, FAX: 6142920603 TEL: 6142929371.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..246  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="60036027R1"  
 /clone\_lib="Arabidopsis developing seed"  
 /tissue\_type="seed"  
 /dev\_stage="5-13 days after flowering"  
 /lab\_host="E.coli"  
 /note="Organ: Developing seed. Vector: pBluescript SK-;  
 Site\_1: EcorI; Site\_2: XhoII"  
 BASE COUNT 66 a 63 c 54 g 63 t  
 ORIGIN  
 Query Match 1.0%; Score 19; DB 10; Length 246;  
 Best Local Similarity 100.0%; Pred. No. 4,2e+02; Mismatches 0; Gaps 0;  
 Matches 19; Conservative 0; Indels 0;  
 QY 1478 TTGATCATTTTTCATCGTA 1496  
 |||||||TTTTTTTTTTT  
 Db 74 TTGATCATTTTTCATCGTA 92  
 RESULT 22 262 bp mRNA EST 13-APR-1999  
 LOCUS AI445399  
 DEFINITION tJ24N08.x1 NCI CGAP Gas4 Homo sapiens CDNA clone IMAGE:2142495 3'  
 similar to gb:M24095 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,  
 A-26(A-10) A\*2601 ALPHA (HUMAN);, mRNA sequence.  
 AI445399  
 ACCESSION AI445399.1 GI:4287940  
 VERSION EST.  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 262)  
 NCBI-CCAP http://www.ncbi.nlm.nih.gov/ncicap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strauberg, Ph.D.  
 Email: ccgaps-remail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 www-bio.llnl.gov/bdrp/image/image.html  
 Insert Length: 962 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 1.  
 Location/Qualifiers  
 1..262  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2142495"  
 /clone\_lib="NCI CGAP Gas4"  
 /tissue\_type="poorly differentiated adenocarcinoma with  
 signet ring cell features"



Query Match 1.0%; Score 19; DB 9; Length 298;  
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 894 TGGGATGATGATTTAAAGG 912  
 Db 214 TGGGATGATGATTTAAAGG 232

RESULT 25  
 A1556340/C 299 bp mRNA linear EST 23-MAR-1999

LOCUS A1556340  
 DEFINITION UI-R-C2P-r9-c-07-0-UI.s1 UI-R-C2P Rattus norvegicus cDNA clone  
 UI-R-C2P-r9-c-07-0-UI 3', mRNA sequence.

ACCESSION A1556340  
 VERSION A1556340.1 GI:4488703

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 299)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: meares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the

oligo-dt track served to identify it as a clone from the normalized

adult muscle library. cDNA library preparation: M. Fatima Bonaldo,

Ph.D. Clone distribution: clones will be available through Research

Genetics (www.resgen.com).

Seq primer: M13 forward.

#### Location/Qualifiers

1. 299

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-C2P-r9-c-07-0-UI"

/clone\_1lb="UI-R-C2P"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C2P

library is a subtracted library derived from the UI-R-C1

library, which is a subtracted library derived from the

UI-R-C0 library. The UI-R-C0 library consisted of a

mixture of individually tagged normalized libraries

constructed from rat placenta, adult lung, brain, liver,

kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day

embryo. The tag is a string of 3-5 nucleotides present

between the Not I site and the oligo-dt track which allows

identification of the library of origin of a clone within

the mixture. The subtracted library (UI-R-C2P) was

constructed as follows: PCR amplified cDNA inserts from

UI-R-C1 clones from which 3' ESTs had been derived was

used as a driver in a hybridization with the UI-R-C1

library in the form of single-stranded circles. The

remaining single-stranded circles (subtracted library) was

purified by hydroxyapatite column chromatography,

converted to double-stranded circles and electroporated

into DH10B bacteria (Life Technologies) to generate the

UI-R-C2P library. This procedure has been previously

described (Bonaldo, Lennon and Soares, Genome Research 6:

BASE COUNT 791-806, 1996)"

ORIGIN

Query Match 1.0%; Score 19; DB 9; Length 299;  
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1842 TTTATGACAAATTCAGATA 1860  
 Db 138 TTTATGACAAATTCAGATA 120

RESULT 26  
 A146527 306 bp mRNA linear EST 03-JUL-1999

LOCUS A146527  
 DEFINITION A146527 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA  
 clone 2810454K18, mRNA sequence.

ACCESSION A146527.1 GI:5350660

VERSION A146527

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 306)

AUTHORS Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,

Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara

A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,

Kikuchi,N., Kojima,Y., Matsuyama,T., Nishitani,H., Oda,H., Owa,C.,

Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara

Y., Suzuki,H., Suzuki,H., Tachino,M., Tomaru,Y., Tomimaga,N.,

Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,

Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.

RIKEN Mouse ESTs

Unpublished (1999)

Contact: Chie Owa

Genome Science Laboratory

RIKEN 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp

Thermotabilization and thermocyclization of thermostable enzymes by

trehalose and its application for the synthesis of full length cDNA

(Proc. Natl. Acad. Sci. U.S.A. 95(12):520-524 (1998))

Transcriptional sequencing: A method for DNA sequencing using RNA

polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))

Please visit our web site (http://genome.rtc.riken.go.jp) for

further details.

#### Location/Qualifiers

1. 306

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="2810454K18"

/clone\_1lb="Mus musculus C57BL/6J 10-11 day embryo"

/sex="mixed"

/dev\_stage="10-11 day embryo"

/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia

) with a modified polylinker; Site 1: Not I; Site 2: Eco

RI; 1st strand cDNA was primed with a Not I - oligo(dT)

primer [5']

TGTTTCACATCGAAGTGGAGCGCGCGGATGTTTTTTTTTTTTTTTTTTTTTTT

T 3'; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pT73 vector.

RNA provided by Dr. Minoru Ko, Wayne State Univ. Library

constructed and normalized by Bento Soares and M. Fatima

Bonaldo."

BASE COUNT 94 a 58 c 68 g 86 t

ORIGIN



```

Query Match      1.0%; Score 19; DB 9; Length 306;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 894 TGGGAATGGATTTTAAAGG 912
|||||
Db 225 TGGGAATGGATTTTAAAGG 243

RESULT 27
AV117891
LOCUS
DEFINITION
AV117891 Mus musculus C57BL/6J 10-day embryo Mus musculus cDNA
clone 2610205N14, mRNA sequence.
ACCESSION
AV117891
VERSION
AV117891.1 GI:5300042
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 319)
AUTHORS
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,
A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,
Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H., Owa, C.,
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara
Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y., Tomimaga, N.,
Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,
Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
TITLE
RIKEN Mouse ESTs
JOURNAL
Unpublished (1999)
COMMENT
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermolabile enzymes by
Thermolabile and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for RNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
source
Location/Qualifiers
1..319
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2610205N14"
/clone_lib="Mus musculus C57BL/6J 10-day embryo"
/sex="mixed"
/dev_stages="10-day embryo"
/notes="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
RI) with a modified polylinker; Site 1: Not I; Site 2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo (dT)
primer [5',
TGTTCACCACTGAGTGGGCGCGCGCGGATCGTTTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
BASE COUNT 93 a 58 c 65 g 103 t
ORIGIN
Query Match      1.0%; Score 19; DB 9; Length 319;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 894 TGGGAATGGATTTTAAAGG 912
|||||
Db 225 TGGGAATGGATTTTAAAGG 243

RESULT 28
AA081640
LOCUS
DEFINITION
AA081640.1 GI:1623698
ACCESSION
AA081640
VERSION
AA081640.1
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 332)
AUTHORS
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chispeili, B.,
Chispe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins
M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E.,
Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28W13 rev2 from Amersham.
FEATURES
source
Location/Qualifiers
1..332
/organism="Homo sapiens"
/db_xref="GDB:3926488"
/db_xref="taxon:9606"
/clone="IMAGE:547972"
/clone_lib="Stratagene neuroepithelium NT2RAMI 937234"
/dev_stages="Ntera-2/RA-MI neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; Cloned unidirectionally.
(Ntera-2/ci.D1) precursor cells induced with Retinoic
Acid for 1 week, followed by 3 weeks in mitotic inhibitors
(Replate #2). Average insert size: 1.1 kb; Uni-ZAP XR
Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 92 a 77 c 78 g 83 t 2 others
ORIGIN
Query Match      1.0%; Score 19; DB 9; Length 332;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCTGTCTCTCTGAATCTT 388
|||||
Db 236 AGCTGTCTCTCTGAATCTT 254

RESULT 29
AI466079
LOCUS
DEFINITION
AI466079 y1 Soares mammary gland NbWMG Mus musculus cDNA clone
IMAGE:1246189 5', mRNA sequence.
ACCESSION
AI466079
VERSION
AI466079.1 GI:4320109
KEYWORDS
EST.
SOURCE
house mouse.

```

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 350)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
Tumor Gene Index  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:659877  
This read is a RESEQUENCE of a previously sequenced mouse clone  
This read has been verified (found to hit its original self in the  
correct orientation)  
Seq primer: -40RP from Gibco  
High quality sequence stop: 333.  
Location/Qualifiers  
1..350  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="IMAGE:1246189"  
/clone\_1lb="Soares\_mammary\_gland\_NbMWG"  
/sex="male"  
/tissue\_type="mammary\_gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Organ: mammary gland; Vector: pRTT3D-Pac (Pharmacia  
) with a modified polylinker. Site 1: Not I; Site 2: Eco  
RI; 1st strand cDNA was primed with a Not I - 01lgo (dT)  
primer [5'  
TGTTCACCAATCTGAGTGGAGCGCGCGAGATGTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pRTT3 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M. Fatima  
Bonaldo."

BASE COUNT 83 a 68 c 86 g 112 t 1 others  
ORIGIN

Query Match 1.0%; Score 19; DB 9; Length 350;  
Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1645 CTTTAATTTTGCATTATT 1663  
|||||  
Db 236 CTTTAATTTTGCATTATT 254

RESULT 30  
LOCUS AM436624 353 bp mRNA linear EST 09-JUL-2000  
DEFINITION 77002 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.  
ACCESSION AM436624  
VERSION AM436624.1 GI:6971930  
KEYWORDS EST.  
SOURCE pig.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
REFERENCE 1 (bases 1 to 353)  
Fahrenkrug, S.C., Fekking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,  
Stone, R.T., Heaton, M.P., Grose, W.M., Bennett, G.A., Laegreid, W.M.  
and Keeler, J.W.  
Design and use of two pooled tissue normalized cDNA libraries for  
EST discovery in swine  
Unpublished (2000)  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA

TITLE  
JOURNAL  
COMMENT

Tel: 402 762 4366  
Fax: 402 762 4390  
Email: [smith@meat.marc.usda.gov](mailto:smith@meat.marc.usda.gov)  
Single pass sequencing. Bases called and trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 20  
and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCCACTACAGAG  
Plate: 32 row: N column: 5  
Seq primer: ATTAGTGACACTATAG.  
Location/Qualifiers  
1..353  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone\_lib="MARC 2P1G"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;  
Library made from pooled tissue from testis, ovary,  
endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 102 a 81 c 75 g 95 t  
ORIGIN

Query Match 1.0%; Score 19; DB 10; Length 353;  
Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 674 TTAGAAATTTCTCTTCAA 692  
|||||  
Db 248 TTAGAAATTTCTCTTCAA 230

RESULT 31  
LOCUS BF364126 356 bp mRNA linear EST 24-NOV-2000  
DEFINITION QV0-NN1022-230600-282-c09 NN1022 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF364126  
VERSION BF364126.1 GI:11326151  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 356)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brito, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
CONTACT: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?i=QV0&c2=QV0-NN1022-230600-282-c09&c3=2000-06-23&t=1>)  
Seq primer: puc 18 forward  
High quality sequence start: 7  
High quality sequence stop: 27.  
Location/Qualifiers  
1..356  
/organism="Homo sapiens"

TITLE  
JOURNAL  
COMMENT

FEATURES  
source

```

/db_xref="taxon:9606"
/clone_lib="NN1022"
/dev_stage="Adult"
/note="Organ: nervous normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORSTES PCR (U.S. Letters Patent application
No. 396,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      111 a      74 c      77 g      94 t
ORIGIN
Query Match
Best Local Similarity 1.0%; Score 19; DB 12; Length 356;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCTGTTCTTCTGAATCTT 388
|||||
Db 151 AGCTGTTCTTCTGAATCTT 169

RESULT 32
AZ537688/c
LOCUS
DEFINITION
AST-DHW(7) Genetrapp PC-3 Human Prostatic Carcinoma Library Homo
sapiens genomic 5', DNA sequence.
ACCESSION
AZ537688
VERSION
AZ537688.1 GI:11115054
KEYWORDS
GSS.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 367)
AUTHORS
Henkel, G., Liyanage, M., Pratt, E., Huang, D., Riley, M., Bernardino, A.,
Durick, K. and Pollok, B.
TITLE
Exon-trap tags from a PC-3 GenomeScreen(TM) Library
JOURNAL
Unpublished (2000)
COMMENT
Contact: Greg Henkel
Gene Expression
Aurora Biosciences Corp.
11010 Torreyana Road, San Diego, CA 92121, USA
Tel: 8584048436
Fax: 8584046719
Email: henkelg@aurorabio.com
Pools of cells were isolated from a GenomeScreen(TM) library. The
library of cells was generated by retroviral integration of a gene
tagging element consisting of: 1) A promoterless beta-lactamase
preceded by a splice acceptor as a reporter for gene expression;
2) A promoter driving neomycin resistance followed by a splice
donor to trap downstream exons. 3' RACE from neomycin gene was
performed using total RNA from isolated pools. Output was shotgun
cloned in pAMP-1 and used to transform DH5-alpha competent
bacteria. 5' ends of reported sequences were immediately preceded
by splice donor from the trapping construct.
Class: exon-trapped.
FEATURES
Location/Qualifiers
source
1..367
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Genetrapp PC-3 Human Prostatic Carcinoma
Library"
/tissue_type="Adenocarcinoma"
/cell_type="Epithelial"
/cell_line="PC-3"
/note="Organ: Prostate; Vector: pAMP-1; 3' RACE of total
RNA from genetrapp pools; shotgun clone in pAMP-1 and used
to transform DH5-alpha competent bacteria."
BASE COUNT      109 a      79 c      96 g      83 t
ORIGIN
Query Match
1.0%; Score 19; DB 17; Length 367;

```

```

Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 609 TGGGCTCAGCACTTCTTTT 627
|||||
Db 62 TGGGCTCAGCACTTCTTTT 44

RESULT 33
BQ528747
LOCUS
DEFINITION
3524_1_37_1_H11.Y.1 3524 - Mature pollen from Sheila McCormick's
lab Zea mays cDNA, mRNA sequence.
ACCESSION
BQ528747
VERSION
BQ528747.1 GI:21390698
KEYWORDS
EST.
SOURCE
Zea mays.
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1 (bases 1 to 376)
AUTHORS
Walbot, V.
TITLE
Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL
Unpublished (1999)
COMMENT
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 3524.1.37.1 row: H column: 11.
FEATURES
Location/Qualifiers
source
1..376
/organism="Zea mays"
/db_xref="B73"
/cultivar="B73"
/cdb_xref="taxon:4577"
/clone_lib="3524 - Mature pollen from Sheila McCormick's
lab"
/tissue_type="pollen"
/dev_stage="mature"
/lab_host="SOLR"
/note="Vector: Stratagene's Uni-Zap XR (pBluescript SK-);
Site 1: EcoRI (5-prime); Site 2: XhoI (3-prime);
Unamplified cDNA library directionally cloned by Rima
Kulikauskas using Stratagene's Uni-Zap system. Inert
sizes ranged from 0.5Kb to 2Kb. 50 microliter aliquot had
338,000 pfu when it was made in Sept, 1995, from oligo
dT-primed poly A+ RNA."
BASE COUNT      100 a      92 c      97 g      87 t
ORIGIN
Query Match
1.0%; Score 19; DB 14; Length 376;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 374 GTTCTTCTGAATCTTTTGC 392
|||||
Db 269 GTTCTTCTGAATCTTTTGC 287

RESULT 34
BJ541842/c
LOCUS
DEFINITION
BJ541842 MFO1SSB cDNA Oryzias latipes cdna clone MFO1SSB036B15 3',
mRNA sequence.
ACCESSION
BJ541842
VERSION
BJ541842.1 GI:22200654
KEYWORDS
EST.
SOURCE
Japanese medaka.
ORGANISM
Oryzias latipes

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.  
 1 (bases 1 to 389)  
 Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.  
 Medaka EST Project in Takeda's lab  
 Unpublished (2001)  
 Contact: Tadasu Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshin@genes.nig.ac.jp.  
 Location/Qualifiers  
 1..389  
 /organism="Oryzias latipes"  
 /db\_xref="taxon:8090"  
 /clone\_lib="MF01SSB036B15"  
 /sex="mixture of female and male"  
 /tissue\_type="whole embryo"  
 /dev\_stage="segmentation stage 20 - 25"  
 /dev\_97 c 104 g 67 t

BASE COUNT 121 a 97 c 104 g 67 t  
 ORIGIN

Query Match 1.0%; Score 19; DB 13; Length 389;  
 Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 849 TTTTAGGTTGCGTTCCT 867  
 |||||  
 Db 143 TTTTAGGTTGCGTTCCT 125

RESULT 35  
 LOCUS T27099 396 bp mRNA linear EST 25-JUL-1996  
 DEFINITION NIBT285H12R Infant brain, LNL array of Dr. M. Soares INIB Homo sapiens CDNA clone LAB285H12 5' end, mRNA sequence.  
 T27099  
 ACCESSION T27099.1 GI:774141  
 VERSION EST.  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 396)  
 Sikeia,U.M.  
 Human cDNAs from infant brain  
 Unpublished (1993)  
 Contact: Sikeia JM  
 Department of Pharmacology  
 University of Colorado Health Sciences Center  
 Box C236, 4200 E. 9th Ave, Denver CO 80262-0236  
 Tel: 3032708637  
 Fax: 3032707097  
 Email: nikki@tally.uchsc.edu  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1..396  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="LAB285H12"  
 /clone\_lib="Infant brain, LNL array of Dr. M. Soares  
 INIB"  
 /note="Vector: lafmid BA; Site\_1: HindIII; Site\_2: NotI; Normalized infant brain CDNA library made by Dr. M. Soares (Columbia University), oligo-dT primed and directionally cloned between HindIII (5') and NotI (3') sites "  
 BASE COUNT 101 a 102 c 104 g 87 t 2 others

ORIGIN  
 Query Match 1.0%; Score 19; DB 14; Length 396;  
 Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 370 AGCTGTTCTTCGATCTT 388  
 |||||  
 Db 344 AGCTGTTCTTCGATCTT 362

RESULT 36  
 LOCUS BJ529530 411 bp mRNA linear EST 09-AUG-2002  
 DEFINITION BJ529530 MF01SSB CDNA Oryzias latipes CDNA clone MF01SSB036B15 5', mRNA sequence.  
 BJ529530  
 ACCESSION BJ529530.1 GI:22188342  
 VERSION EST.  
 KEYWORDS Japanese medaka.  
 SOURCE Oryzias latipes  
 ORGANISM Oryzias latipes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.  
 1 (bases 1 to 411)  
 Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.  
 Medaka EST Project in Takeda's lab  
 Unpublished (2001)  
 Contact: Tadasu Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshin@genes.nig.ac.jp.  
 Location/Qualifiers  
 1..411  
 /organism="Oryzias latipes"  
 /strain="Hd-rR"  
 /db\_xref="taxon:8090"  
 /clone\_lib="MF01SSB036B15"  
 /clone\_lib="MF01SSB CDNA"  
 /sex="mixture of female and male"  
 /tissue\_type="whole embryo"  
 /dev\_stage="segmentation stage 20 - 25"  
 /dev\_104 c 99 g 128 t

BASE COUNT 80 a 104 c 99 g 128 t  
 ORIGIN

Query Match 1.0%; Score 19; DB 13; Length 411;  
 Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 849 TTTTAGGTTGCGTTCCT 867  
 |||||  
 Db 249 TTTTAGGTTGCGTTCCT 267

RESULT 37  
 LOCUS AA822992 419 bp mRNA linear EST 17-FEB-1998  
 DEFINITION AA822992 vW39607.r1 Soares mammary\_gland\_NDMMG Mus musculus CDNA clone IMAGE:1246189 5', mRNA sequence.  
 AA822992  
 ACCESSION AA822992.1 GI:2892860  
 VERSION EST.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 419)  
 Marx,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kueba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterson, R.

# TITLE JOURNAL COMMENT

The WashU-HHMI Mouse EST Project  
Unpublished (1996)  
Contact: Martha M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (infoimage.llnl.gov) for further information.

Seq primer: -28m13 rev2 ET from Amersham.

## FEATURES source

```
1. .419
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1246189"
/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stages="4 weeks"
/lab_host="DH10B"
/notes="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTACCAATCTGAAGTGGAGCGCCGCGAATGTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
```

BASE COUNT 95 a 86 c 98 g 140 t

Query Match 1.0%; Score 19; DB 9; Length 419;  
Best Local Similarity 100.0%; Pred. NO. 4.5e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1645 CTTTAATTTTGCATTATT 1663  
|||||  
Db 236 CTTTAATTTTGCATTATT 254

## RESULT 38 AZ290933

LOCUS AZ290933 430 bp DNA linear GSS 27-JUL-2000  
DEFINITION RPCI-23-103A23.TJ RPCI-23 Mus musculus genomic clone RPCI-23-103A23  
; DNA sequence.

ACCESSION AZ290933  
VERSION AZ290933.1 GI:9532719  
KEYWORDS GSS.  
SOURCE house mouse.

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 430)  
Zhao, S., Niemman, W., Feldblum, T., Malek, J., Shateman, S., Akinret  
B., Levins, M., McGann, S., Teegaye, G., Geer, K., Krol, M., de Jong, P.  
and Fraser, C.M.

## TITLE

Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)  
or from Resea ch Genetics (info@resgen.com). BAC end page:  
[http://www.tigr.org/tdb/bac\\_end/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_end/mouse/bac_end_intro.html)  
Plate: 103 row: A column: 23  
Seq primer: SP6

Class: BAC ends.  
Location/Qualifiers

## source

```
1. .430
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-103A23"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:  
ECORI; Site 2: EORI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EORI and EORI Methyllase. Size  
selected DNA was cloned into the pBACe3.6 vector at the  
ECORI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies). "
```

BASE COUNT 139 a 66 c 91 g 134 t

## ORIGIN

Query Match 1.0%; Score 19; DB 17; Length 430;  
Best Local Similarity 100.0%; Pred. NO. 4.5e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1712 CCACCTCTGTGTGTGTTTA 1730

Db 144 CCACCTCTGTGTGTGTTTA 162

## RESULT 39

## AQ359339/c

LOCUS AQ359339 430 bp DNA linear GSS 06-MAR-1999  
DEFINITION HS\_5030\_B2\_D09\_SP6E RPCI11 Human Male BAC Library Homo sapiens  
genomic clone Plate=606 Col=18 Row=H, DNA sequence.

ACCESSION AQ359339

VERSION AQ359339.1 GI:4208215

KEYWORDS GSS.

SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

## AUTHORS

Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., and  
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and  
Hood, L.

Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

## JOURNAL

## MEDLINE

## COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 606 row: H column: 18

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 430.

## FEATURES

Location/Qualifiers

```

source
1. .430
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=606 Col=18 Row=H"
/clone_lib="RPC111 Human Male BAC Library"
/sex="Male"
/cell_type="lymphocytes"
/note="Vector: pBAC3.6; RPC111 Human Male BAC Library"

BASE COUNT      113 a      91 c      93 g      132 t      1 others
ORIGIN
Query Match      1.0%; Score 19; DB 17; Length 430;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1151 ATTGTTGTCAGAACAGGC 1169
|||||
Db 245 ATTGTTGTCAGAACAGGC 227

RESULT 40
BH864257 447 bp DNA linear GSS 05-AUG-2002
LOCUS     SALK_095658 Arabidopsis thaliana TDNA insertion lines Arabidopsis
DEFINITION thaliana genomic clone SALK_095658, DNA sequence.
ACCESSION BH864257 GI:22100155
VERSION   BH864257.1
KEYWORDS  thale cress.
SOURCE    Arabidopsis thaliana
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 447)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
,Zimmerman,J. and Ecker,J.R.
A Sequence-indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA. This sequence lies within an annotated exon of Atlg23010.
Class: TDNA tagged.
Location/Qualifiers
1. .447
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_095658"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/cdna_protocols.html"

BASE COUNT      104 a      104 c      114 g      125 t
ORIGIN
Query Match      1.0%; Score 19; DB 17; Length 447;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1478 TTGATCATTTTGATCGTA 1496
|||||
Db 163 TTGATCATTTTGATCGTA 145

RESULT 41
AV968816/c 455 bp mRNA linear EST 14-MAR-2002
LOCUS     AV968816 Nori Satoh unpublished cDNA library, larva Ciona
DEFINITION intestinalis cDNA clone c1v16p19 5', mRNA sequence.
ACCESSION AV968816 GI:19456580
VERSION   AV968816.1
KEYWORDS  Ciona intestinalis.
SOURCE    Ciona intestinalis.
ORGANISM  Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.
1 (bases 1 to 455)
Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.
Expressed genes in Ciona intestinalis
Unpublished (2000)
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1. .455
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="c1v16p19"
/clone_lib="Nori Satoh unpublished cDNA library, larva"
/tissue_type="whole animal"
/dev_stage="larva"
/note="Vector: pBluescript SK"

BASE COUNT      149 a      77 c      113 g      113 t      3 others
ORIGIN
Query Match      1.0%; Score 19; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 210 CGATGCTTTTGTCCCGA 228
|||||
Db 354 CGATGCTTTTGTCCCGA 336

RESULT 42
AA798200 459 bp mRNA linear EST 10-FEB-1998
LOCUS     ub60e03.r1 Soares mammary_gland NMLMG Mus musculus cDNA clone
DEFINITION IMAGE:1382140 5' similar to TR:O10384 O10384 GAG POLYPEPTIDEIN. [1]
ACCESSION AA798200 GI:2861155
VERSION   AA798200.1
KEYWORDS  Mus musculus.
SOURCE    house mouse.
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 459)
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

```

Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:904608  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 451.  
Location/Qualifiers  
1..459  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1382140"  
/clone\_lib="Soares mammary\_gland\_NMLMG"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"  
/note="vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT, 137 a 106 c 92 g 124 t  
ORIGIN

Query Match 1.0%; Score 19; DB 9; Length 459;  
Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 894 TCGGAATCGATTTAAGG 912  
|||||  
Db 421 TCGGAATCGATTTAAGG 439

RESULT 43  
LOCUS AQ698248 459 bp DNA linear GSS 06-JUL-1999  
DEFINITION HS 5550 B2 E04 SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1126 Col=8 Row=J. DNA sequence.  
ACCESSION AQ698248  
VERSION AQ698248.1 GI:5388496  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 459)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
MEDLINE 93380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallaceu.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end web Server: http://www.htsc.washington.edu  
Plate: 1126 row: J column: 8  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 459.  
Location/Qualifiers

FEATURES  
source  
1..459  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=1126 Col=8 Row=J"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="Male"  
/note="vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"  
BASE COUNT 118 a 81 c 104 g 156 t  
ORIGIN

Query Match 1.0%; Score 19; DB 17; Length 459;  
Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 AGAGGAATTTTATGTCT 529  
|||||  
Db 310 AGAGGAATTTTATGTCT 328

RESULT 44  
LOCUS A1159020 461 bp mRNA linear EST 02-OCT-1998  
DEFINITION V286f10.r1 Soares\_mammary\_gland\_NbMMG Mus musculus cDNA clone IMAGE:1333387 5', mRNA sequence.  
ACCESSION A1159020  
VERSION A1159020.1 GI:3692202  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 461)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geiscl,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Scheilenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.  
TITLE The WashU-HMMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:692931  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 436.  
Location/Qualifiers  
1..461  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1333387"  
/clone\_lib="Soares\_mammary\_gland\_NbMMG"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5',  
TGTTACCAATCTGAATGGAGCGGCCGAATGTTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pRT73 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M. Fatima  
Bonaldo."

BASE COUNT 127 a 78 c 136 g 120 t

BASE COUNT 193 a 85 c 74 g 129 t

Query Match 1.0%; Score 19; DB 9; Length 461;  
Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 1.0%; Score 19; DB 17; Length 481;  
Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1645 CTTTAAATTTGCAATTAAT 1663  
Db 381 CTTTAAATTTGCAATTAAT 399

Oy 1460 TATTCATTAACAACTACTT 1478  
Db 458 TATTCATTAACAACTACTT 476

# RESULT 45

AZ463779

481 bp DNA linear GSS 04-OCT-2000

LOCUS 1M0272A19R Mouse 10kb plasmid UGCM library Mus musculus genomic  
DEFINITION clone UGCM0272A19 R, DNA sequence.

ACCESSION

AZ463779

VERSION

AZ463779.1 GI:10621904

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

# REFERENCE

1 (bases 1 to 481)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
'M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.  
and Wright, D., Weiss, R.

# TITLE

Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

# JOURNAL

Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center

COMMENT

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA

Tel: 801 585 5606  
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000

Std Error: 0.00

Plate: 0272

row: A column: 19

Seq primer: CACACAGAAACGCTATGACC

Class: plasmid ends

High quality sequence stop: 481.

Location/Qualifiers

# FEATURES

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481

Location/Qualifiers

1. 481



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 10:34:31, Search time 73 Seconds  
(without alignments)  
10097.799 Million cell updates/sec

Title: US-09-868-987-1  
Perfect score: 1864  
Sequence: 1 atgacttcgcataattgtc.....atgacaaattcagataatgc 1864

Scoring table: OIIGO NUC  
Gapop 60.0, Gapext 60.0

Searched: 355320 seqs, 197730502 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 710640

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1864	100.0	1864	10	US-09-886-468-1
2	20	1.1	448	10	US-09-983-965-4080
C 3	19	1.0	439	10	US-09-864-761-10333
C 4	19	1.0	445	10	US-09-864-761-26966
C 5	19	1.0	469	10	US-09-864-761-32700
C 6	19	1.0	561	10	US-09-864-761-16209
7	19	1.0	1150	10	US-09-822-849A-374
8	19	1.0	1707	9	US-09-938-842A-186
9	19	1.0	2000	9	US-09-938-842A-4668
C 10	19	1.0	3470	12	US-10-044-090-516
11	18	1.0	366	10	US-09-878-574-2011
12	18	1.0	392	10	US-09-924-035A-470
C 13	18	1.0	421	10	US-09-983-965-2819
C 14	18	1.0	459	10	US-09-920-300A-17
C 15	18	1.0	459	12	US-10-033-528-17
C 16	18	1.0	478	10	US-09-770-444-3
17	18	1.0	613	10	US-09-770-149-852
18	18	1.0	713	10	US-09-764-847-1360
19	18	1.0	755	9	US-09-938-842A-5222
					Sequence 1, Appli
					Sequence 4080, Ap
					Sequence 10333, A
					Sequence 26966, A
					Sequence 32700, A
					Sequence 16209, A
					Sequence 374, App
					Sequence 186, App
					Sequence 4668, Ap
					Sequence 516, App
					Sequence 470, App
					Sequence 2819, Ap
					Sequence 17, Appl
					Sequence 3, Appli
					Sequence 852, App
					Sequence 1360, Ap
					Sequence 5222, Ap

Sequence 715, App  
Sequence 1361, Ap  
Sequence 2071, Ap  
Sequence 2072, Ap  
Sequence 3, Appli  
Sequence 26319, A  
Sequence 1387, Ap  
Sequence 1387, Ap  
Sequence 452, App  
Sequence 6153, Ap  
Sequence 3885, Ap  
Sequence 5590, Ap  
Sequence 1850, Ap  
Sequence 10630, A  
Sequence 67, Appli  
Sequence 7444, Ap  
Sequence 65, Appli  
Sequence 157, App  
Sequence 9, Appli  
Sequence 161, App  
Sequence 5070, Ap  
Sequence 9546, Ap  
Sequence 116, App  
Sequence 173, App  
Sequence 1, Appli  
Sequence 3, Appli  
Sequence 4722, Ap  
Sequence 8636, Ap  
Sequence 6, Appli  
Sequence 1582, Ap  
Sequence 1584, Ap  
Sequence 1583, Ap  
Sequence 1, Appli  
Sequence 46, Appli  
Sequence 46, Appli  
Sequence 2184, Ap  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 99, Appli  
Sequence 99127, A  
Sequence 172, App  
Sequence 32118, A  
Sequence 3542, Ap  
Sequence 967, App  
Sequence 1319, Ap  
Sequence 13252, A  
Sequence 6487, App  
Sequence 536, App  
Sequence 1539, Ap  
Sequence 167, App  
Sequence 112, App  
Sequence 1194, Ap  
Sequence 222, App  
Sequence 436, App  
Sequence 267, App  
Sequence 1863, Ap  
Sequence 11588, A  
Sequence 65, Appli  
Sequence 489, App  
Sequence 8773, Ap  
Sequence 4878, Ap  
Sequence 14446, A  
Sequence 1078, Ap  
Sequence 9849, Ap  
Sequence 4026, Ap  
Sequence 7924, Ap  
Sequence 264, App  
Sequence 15066, A  
Sequence 1051, Ap







531	15	0.8	750	10	US-09-974-300-6310	Sequence 6310, Ap	c 604	15	0.8	1323	10	US-09-778-320-38	Sequence 38, Appl
532	15	0.8	753	10	US-09-974-300-7239	Sequence 7239, Ap	c 605	15	0.8	1323	10	US-09-910-689-38	Sequence 38, Appl
533	15	0.8	768	10	US-09-910-943-187	Sequence 187, Ap	c 606	15	0.8	1323	12	US-10-010-742-38	Sequence 38, Appl
534	15	0.8	771	10	US-09-974-300-2926	Sequence 2926, Ap	607	15	0.8	1325	10	US-09-765-205-1	Sequence 1, Appl
535	15	0.8	773	9	US-09-771-035A-52	Sequence 52, Appl	608	15	0.8	1325	9	US-09-978-295A-321	Sequence 321, App
536	15	0.8	789	9	US-09-813-453A-31	Sequence 31, Appl	609	15	0.8	1333	9	US-09-978-697-321	Sequence 321, App
537	15	0.8	813	10	US-09-770-445-763	Sequence 763, App	610	15	0.8	1333	9	US-09-978-842A-321	Sequence 321, App
538	15	0.8	836	10	US-09-770-445-675	Sequence 675, App	611	15	0.8	1344	9	US-09-938-842A-2666	Sequence 2666, Ap
539	15	0.8	852	10	US-09-770-445-623	Sequence 623, App	612	15	0.8	1345	9	US-09-938-842A-4296	Sequence 4296, Ap
540	15	0.8	854	10	US-09-770-445-618	Sequence 618, App	613	15	0.8	1345	9	US-09-938-842A-5125	Sequence 5125, Ap
541	15	0.8	877	9	US-09-938-842A-5040	Sequence 5040, Ap	614	15	0.8	1350	9	US-09-895-913A-291	Sequence 291, App
542	15	0.8	891	10	US-09-815-242-7115	Sequence 7115, Ap	c 615	15	0.8	1356	10	US-09-815-242-6624	Sequence 6624, Ap
543	15	0.8	893	10	US-09-764-878-277	Sequence 277, App	616	15	0.8	1362	9	US-09-974-298-73	Sequence 73, Appl
544	15	0.8	893	10	US-09-764-878-277	Sequence 277, App	617	15	0.8	1362	12	US-10-044-090-556	Sequence 556, App
545	15	0.8	894	10	US-09-764-860-1140	Sequence 1140, Ap	618	15	0.8	1371	10	US-09-801-368-203	Sequence 203, App
546	15	0.8	894	10	US-09-070-927A-683	Sequence 683, App	619	15	0.8	1377	10	US-09-801-368-167	Sequence 167, App
547	15	0.8	912	9	US-09-764-868-571	Sequence 571, App	620	15	0.8	1386	10	US-09-974-300-1151	Sequence 1151, Ap
548	15	0.8	912	9	US-09-764-868-581	Sequence 581, App	621	15	0.8	1389	9	US-09-938-842A-340	Sequence 340, App
549	15	0.8	930	10	US-09-815-242-7193	Sequence 7193, Ap	c 622	15	0.8	1389	9	US-09-938-842A-113	Sequence 113, App
550	15	0.8	936	10	US-09-815-242-4764	Sequence 4764, Ap	c 623	15	0.8	1391	12	US-10-044-477-2	Sequence 2, Appl
551	15	0.8	950	10	US-09-886-468-12	Sequence 12, Appl	624	15	0.8	1404	10	US-09-729-835-34	Sequence 34, Appl
552	15	0.8	960	9	US-09-938-842A-1228	Sequence 1228, Ap	625	15	0.8	1428	10	US-09-741-669-265	Sequence 265, App
553	15	0.8	972	10	US-09-815-242-6142	Sequence 6142, Ap	626	15	0.8	1431	10	US-09-774-414-2	Sequence 2, Appl
554	15	0.8	978	10	US-09-815-242-3800	Sequence 3800, Ap	627	15	0.8	1440	9	US-09-938-842A-80	Sequence 80, Appl
555	15	0.8	978	10	US-09-815-242-8849	Sequence 8849, Ap	628	15	0.8	1462	10	US-09-220-091-6	Sequence 6, Appl
556	15	0.8	999	10	US-09-815-242-8974	Sequence 8974, Ap	629	15	0.8	1467	9	US-09-938-842A-970	Sequence 970, App
557	15	0.8	1003	9	US-09-938-842A-4517	Sequence 4517, Ap	630	15	0.8	1470	10	US-09-841-132-410	Sequence 410, App
558	15	0.8	1034	9	US-09-736-457-1654	Sequence 1654, Ap	631	15	0.8	1515	10	US	







C 969	14	0.8	62	10	US-09-878-574-9009	Sequence 9009, Ap
C 970	14	0.8	75	10	US-09-864-761-24218	Sequence 24218, A
C 971	14	0.8	76	10	US-09-864-761-23727	Sequence 23727, A
C 972	14	0.8	80	10	US-09-864-761-28939	Sequence 28939, A
C 973	14	0.8	91	9	US-10-035-098-24	Sequence 24, Appl
C 974	14	0.8	92	12	US-10-029-913-3	Sequence 3, Appl
C 975	14	0.8	97	9	US-10-035-098-23	Sequence 23, Appl
C 976	14	0.8	100	10	US-09-864-761-31448	Sequence 31448, A
C 977	14	0.8	101	10	US-09-867-701-2914	Sequence 2914, Ap
C 978	14	0.8	102	10	US-09-864-761-32561	Sequence 32561, A
C 979	14	0.8	103	9	US-10-046-935-1608	Sequence 1608, Ap
C 980	14	0.8	103	9	US-09-878-178-1608	Sequence 1608, Ap
C 981	14	0.8	106	10	US-09-969-373-1342	Sequence 1342, Ap
C 982	14	0.8	108	10	US-09-864-761-22137	Sequence 22137, A
C 983	14	0.8	111	10	US-09-864-761-23456	Sequence 23456, A
C 984	14	0.8	114	10	US-09-815-242-7020	Sequence 7020, Ap
C 985	14	0.8	114	10	US-09-867-701-1549	Sequence 1549, Ap
C 986	14	0.8	118	10	US-09-864-761-16867	Sequence 16867, A
C 987	14	0.8	124	10	US-09-864-761-30335	Sequence 30335, A
C 988	14	0.8	127	10	US-09-864-761-28853	Sequence 28853, A
C 989	14	0.8	129	10	US-09-923-876-5782	Sequence 5782, Ap
C 990	14	0.8	129	10	US-09-864-761-19910	Sequence 19910, A
C 991	14	0.8	131	10	US-09-867-701-2673	Sequence 2673, Ap
C 992	14	0.8	132	10	US-09-864-761-26847	Sequence 26847, A
C 993	14	0.8	136	10	US-09-867-701-5102	Sequence 5102, Ap
C 994	14	0.8	138	10	US-09-974-300-7573	Sequence 7573, Ap
C 995	14	0.8	141	10	US-09-783-590-9503	Sequence 9503, Ap
C 996	14	0.8	144	10	US-09-770-696-5525	Sequence 5525, App
C 997	14	0.8	148	10	US-09-864-761-21728	Sequence 21728, A
C 998	14	0.8	148	10	US-09-864-761-17911	Sequence 17911, A
C 999	14	0.8	149	10	US-09-864-761-30530	Sequence 30530, A
1000	14	0.8	149	10	US-09-969-373-316	Sequence 316, App

## ALIGNMENTS

```

RESULT 1
US-09-886-468-1
; Sequence 1, Application US/09886468
; Patent No. US20020037293A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pasteur Limited
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
; FILE REFERENCE: 77813-5
; CURRENT APPLICATION NUMBER: US/09/886,468
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/113,280
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,281
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,282
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,283
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,284
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,285
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,385
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114,050
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,056
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,057
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,058
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,059
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,061
; PRIOR FILING DATE: 1998-12-28

```

NUMBER OF SEQ ID NOS: 26									
SOFTWARE: PatentIn Ver. 2.0									
SEQ ID NO 1									
LENGTH: 1864									
TYPE: DNA									
ORGANISM: Chlamydia pneumoniae									
FEATURE:									
NAME/KEY: CDS									
LOCATION: (101)..(1756)									
US-09-886-468-1									
Query Match 100.0%; Score 1864; DB 10; Length 1864;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1864; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	ATGACCTCCGATATTTGTCAGAGGGGATCAGCGGCACTGTAATGACAAATTTCTGC	60						
Db	1	ATGACCTCCGATATTTGTCAGAGGGGATCAGCGGCACTGTAATGACAAATTTCTGC	60						
Qy	61	AAACCGTGATGGCGTATGCTGTATGATTTGACCGTTATATGTCAGACGCCATTCTT	120						
Db	61	AAACCGTGATGGCGTATGCTGTATGATTTGACCGTTATATGTCAGACGCCATTCTT	120						
Qy	121	AAACGTCCTGATGAAATCATGCCAGTCTCAGGGAATTTACCCACGTAAGTGA	180						
Db	121	AAACGTCCTGATGAAATCATGCCAGTCTCAGGGAATTTACCCACGTAAGTGA	180						
Qy	181	CAAACTGCGCTCAGATTTAAATCTGAGCGATGCTTTTGTCCGAGGTTCTCAGTA	240						
Db	181	CAAACTGCGCTCAGATTTAAATCTGAGCGATGCTTTTGTCCGAGGTTCTCAGTA	240						
Qy	241	AGAGAGATCTCTTCTGATCTTGGGAAAAACAATGACAAAGCATATCTCAGCAG	300						
Db	241	AGAGAGATCTCTTCTGATCTTGGGAAAAACAATGACAAAGCATATCTCAGCAG	300						
Qy	301	CTGTGCTTGGCAATGCTTATTTTGTGATGAGCGTATATTAATTTGAGGCGTCA	360						
Db	301	CTGTGCTTGGCAATGCTTATTTTGTGATGAGCGTATATTAATTTGAGGCGTCA	360						
Qy	361	CGCTTGGGAGCGTCTTCTGATCTTGGGAAAAACAATGACAAAGCATATCTCAG	420						
Db	361	CGCTTGGGAGCGTCTTCTGATCTTGGGAAAAACAATGACAAAGCATATCTCAG	420						
Qy	421	TGGCCACTCAGCTTCTGAGACTCGCTGGATGTTCTTGTCTATGAGGATGGCCCTA	480						
Db	421	TGGCCACTCAGCTTCTGAGACTCGCTGGATGTTCTTGTCTATGAGGATGGCCCTA	480						
Qy	481	TGCAATGTTCTGTATTTGAAAGATCCGAGAGAAATTTTATGTCCTCAAGCTTAA	540						
Db	481	TGCAATGTTCTGTATTTGAAAGATCCGAGAGAAATTTTATGTCCTCAAGCTTAA	540						
Qy	541	AAATCTGTAGAAAAGATATCAAGGCTTTGGAGCCATTTTGAATCTTAAGTGA	600						
Db	541	AAATCTGTAGAAAAGATATCAAGGCTTTGGAGCCATTTTGAATCTTAAGTGA	600						
Qy	601	TACAGATTTGGCTCAGCACTTCTTCTTCCATGATTAAGGCCCTATTAAGGTTTC	660						
Db	601	TACAGATTTGGCTCAGCACTTCTTCTTCCATGATTAAGGCCCTATTAAGGTTTC	660						
Qy	661	TTTGACATGATTTTGAATTTCTCTCAATGTTTACGGCTTTTCAATGCTAAAT	720						
Db	661	TTTGACATGATTTTGAATTTCTCTCAATGTTTACGGCTTTTCAATGCTAAAT	720						
Qy	721	TTTCTTCATGCTGTGATGATTAAGCCCAACATACAGTTGATATGATTAAGTT	780						
Db	721	TTTCTTCATGCTGTGATGATTAAGCCCAACATACAGTTGATATGATTAAGTT	780						
Qy	781	CGTGGGATTAAGCATGATTTCTTGAGAGATGCAAAAATTTGGGCTGTTCTGAAG	840						
Db	781	CGTGGGATTAAGCATGATTTCTTGAGAGATGCAAAAATTTGGGCTGTTCTGAAG	840						
Qy	841	TGTTTCTTTAGTTGGCTGCTCGGGGTTTGGAGCTGGAATTCGTTTGGGAAT	900						
Db	841	TGTTTCTTTAGTTGGCTGCTCGGGGTTTGGAGCTGGAATTCGTTTGGGAAT	900						



Db 841 TGTCTTTCTTTTAGTGTGCTCTCGGTTTGGAGCCTGGAATTCGTTTGGGAAT 900  
Qy 901 GGAATTTAAAGAGGGTATGCTTTTACCTTTAAATCAAAAGAGCATGGCATCAGCGATGT 960  
Db 901 GGAATTTAAAGAGGGTATGCTTTTACCTTTAAATCAAAAGAGCATGGCATCAGCGATGT 960  
Qy 961 TGCTCAAAATGGTGGCAAAAGTTGTGCATAAACTACAGGAAGCTGCTTTCTTTAGAGA 1020  
Db 961 TGCTCAAAATGGTGGCAAAAGTTGTGCATAAACTACAGGAAGCTGCTTTCTTTAGAGA 1020  
Qy 1021 CTTCGGTATTCAAAATTTGGATCTTCAGAAAAAGATCAAAATCTATTATTAGTATAAAGC 1080  
Db 1021 CTTCGGTATTCAAAATTTGGATCTTCAGAAAAAGATCAAAATCTATTATTAGTATAAAGC 1080  
Qy 1081 TTTAAGCTACTAAGACAGATACGAGCCTCTCTCTAAATTAACGATCATGAGCTGGCG 1140  
Db 1081 TTTAAGCTACTAAGACAGATACGAGCCTCTCTCTAAATTAACGATCATGAGCTGGCG 1140  
Qy 1141 TTATTGTGGGATTGTTGTACAGAAACAGCGCTAGATTCTCTACGGAAACTCTAAGCGAAA 1200  
Db 1141 TTATTGTGGGATTGTTGTACAGAAACAGCGCTAGATTCTCTACGGAAACTCTAAGCGAAA 1200  
Qy 1201 CGCAAAATTTTGGTCAAAAGTTAAGCAGCAAACTATCGAAGAAAATGCGTTATCAGCGCAC 1260  
Db 1201 CGCAAAATTTTGGTCAAAAGTTAAGCAGCAAACTATCGAAGAAAATGCGTTATCAGCGCAC 1260  
Qy 1261 CATCGGCTTTTAGAGCTTTGGCAATCATCTTGCTATGATGAGTTGCGCTTTGAATG 1320  
Db 1261 CATCGGCTTTTAGAGCTTTGGCAATCATCTTGCTATGATGAGTTGCGCTTTGAATG 1320  
Qy 1321 GCAATATGCTTTCAAGTGGCGTATGCGCTTTAAATCATGACCTTTTGGCTACCTGTGCGAT 1380  
Db 1321 GCAATATGCTTTCAAGTGGCGTATGCGCTTTAAATCATGACCTTTTGGCTACCTGTGCGAT 1380  
Qy 1381 CTGTGTTTATAGCACATTTCTTTTCAAGAAAATTCAAATAGATTTTGCAGCCATTGGTGC 1440  
Db 1381 CTGTGTTTATAGCACATTTCTTTTCAAGAAAATTCAAATAGATTTTGCAGCCATTGGTGC 1440  
Qy 1441 TTTAATGACTGTTATGGGGTATTCATTAACAAATCTTTGATCACTTTTGTATCGTATTGC 1500  
Db 1441 TTTAATGACTGTTATGGGGTATTCATTAACAAATCTTTGATCACTTTTGTATCGTATTGC 1500  
Qy 1501 TGAAGATCGCAAGCGAACCTGTTTACCCCTATGATGCTTTTAGTAAATGATGCCCTTCA 1560  
Db 1501 TGAAGATCGCAAGCGAACCTGTTTACCCCTATGATGCTTTTAGTAAATGATGCCCTTCA 1560  
Qy 1561 AAAGACGTTTCAGCCGACGGTAAATGACAAACAGCTACAATCTATCAGTTTGTGTTAATGCT 1620  
Db 1561 AAAGACGTTTCAGCCGACGGTAAATGACAAACAGCTACAATCTATCAGTTTGTGTTAATGCT 1620  
Qy 1621 TTTGTTTATAGCGGCTCCTCTGCTTTTAAATTTTGCATTTTATTAAGCATAGGATTC 1680  
Db 1621 TTTGTTTATAGCGGCTCCTCTGCTTTTAAATTTTGCATTTTATTAAGCATAGGATTC 1680  
Qy 1681 TCTAGAACTTTATGCTCTCTTTATTTGACCAACCTCTGTTGTTTATGCTCGTAA 1740  
Db 1681 TCTAGAACTTTATGCTCTCTTTATTTGACCAACCTCTGTTGTTTATGCTCGTAA 1740  
Qy 1741 AGAAAATCGCTCAAAATAGTACCGTTAACTTAATCTAACCTGTAGCAATATAAAATC 1800  
Db 1741 AGAAAATCGCTCAAAATAGTACCGTTAACTTAATCTAACCTGTAGCAATATAAAATC 1800  
Qy 1801 TCCTTTGGGACTTTAGTCCCAAGGCCCTGTGGTATTAAATTTATGACAAATTCAGATA 1860  
Db 1801 TCCTTTGGGACTTTAGTCCCAAGGCCCTGTGGTATTAAATTTATGACAAATTCAGATA 1860  
Qy 1861 ATGC 1864  
Db 1861 ATGC 1864

RESULT 2  
US-09-983-965-4080

; Sequence 4080, Application US/09983965  
; Patent No. US20020137160A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Mengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
; FILE REFERENCE: 37-21(10297)C  
; CURRENT APPLICATION NUMBER: US/09/983,965  
; PRIOR FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: US 09/465,231  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: US 60/113,678  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 5912  
; SEQ ID NO 4080  
; LENGTH: 448  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 56-LIB3058-005-Q1-K1-F8  
US-09-983-965-4080  
  
Query Match 1.1%; Score 20; DB 10; Length 448;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1641 CTGCTCTTAATTTTGCATTT 1660  
Db 357 CTGCTCTTAATTTTGCATTT 376  
  
RESULT 3  
US-09-864-761-10333/C  
; Sequence 10333, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aesmica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10333
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC012460.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
US-09-864-761-10333

Query Match          1.0%; Score 19; DB 10; Length 439;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 AGCTGTTCTTGATCTT 388
DB 412 AGCTGTTCTTGATCTT 394

RESULT 4
US-09-864-761-26966/c
; Sequence 26966, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 26966
; LENGTH: 445
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC012460.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
; OTHER INFORMATION: EST HUMAN HIT: BE264998.1, EVALU0 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P34384, EVALU0 1.50e-02
; OTHER INFORMATION: NT HIT: AE002179.2, EVALU0 9.40e-01
US-09-864-761-26966

Query Match          1.0%; Score 19; DB 10; Length 445;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 AGCTGTTCTTGATCTT 388
DB 286 AGCTGTTCTTGATCTT 268

RESULT 5
US-09-864-761-32700/c
; Sequence 32700, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32700
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC012664.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3
; OTHER INFORMATION: SWISSPROT HIT: P34364, EVALU2 = 0.08
; OTHER INFORMATION: NT HIT: AE002179.2, EVALU1 = 1.20e+00
; OTHER INFORMATION: EST HUMAN HIT: BE264998.1, EVALU0 = 0.00e+00
; US-09-964-761-32700

```

```
Query Match      1.0%; Score 19; DB 10; Length 469;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 370 AGCTGTTCTTCTGAATCTT 388  
|||  
Dy 310 AGCTGTTCTTCTGAATCTT 292

Db 310 AGCTGTTCTTCTGAATCTT 292

## RESULT 6

```

RES0116
US-09-864-761-16209/c
; Sequence 16209, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30

```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SEQ ID NO 16209
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC012664.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3
US-09-864-761-16209

```

```
Query Match      1.0%; Score 19; DB 10; Length 561;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0 Indels
```

QY 370 AGCTGTTCTTCTGAATCTT 388

Db 356 AGCTGTTCTTCTGAATCTT 338

## RESULT 7

```

US-09-822-849A-374
; Sequence 374, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fecthel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 374
; LENGTH: 1150
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-374

```

```
Query Match      1.0%; Score 19; DB 10; Length 1150;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 19; Conservative 0; Mismatches 0; Indels
```

QY 370 AGCTGTTCTTCTGAATCTT 388

337 AGCTGTTCTTCTGAATCTT 355

## RESULT 8

US-09-938-842A-186  
; Sequence 186, Application US/09938842A  
; Patent No. US20020160378A1

```

; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 186
; LENGTH: 1707
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-186

Query Match 1.0%; Score 19; DB 9; Length 1707;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1478 TTGATCATTTTTCATCGTA 1496
Db 673 TTGATCATTTTTCATCGTA 691

RESULT 9
US-09-938-842A-4668
; Sequence 4668, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4668
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-4668

Query Match 1.0%; Score 19; DB 9; Length 2000;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1395 ATTCTTTTGAAGAAAT 1413
Db 494 ATTCTTTTGAAGAAAT 512

RESULT 10
US-10-044-090-516/c
; Sequence 516, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 516
; LENGTH: 3470
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1994709CB1
; US-10-044-090-516

Query Match 1.0%; Score 19; DB 12; Length 3470;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1280 TTGGCAATCATCTTGCTCT 1298
Db 1712 TTGGCAATCATCTTGCTCT 1694

RESULT 11
US-09-878-574-2011
; Sequence 2011, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 2011
; LENGTH: 366
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-030-01-B1-C1
; US-09-878-574-2011

Query Match 1.0%; Score 18; DB 10; Length 366;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1785 TAGCAATATTAATAATCTC 1802
Db 7 TAGCAATATTAATAATCTC 24

RESULT 12
US-09-924-035A-470
; Sequence 470, Application US/09924035A
; Patent No. US20020142319A1
; GENERAL INFORMATION:
; APPLICANT: Grilach, Jyn
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2011US
; CURRENT APPLICATION NUMBER: US/09/924,035A
; CURRENT FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/148,784
; PRIOR FILING DATE: 1999-08-13
; NUMBER OF SEQ ID NOS: 900
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 470
```

```

; LENGTH: 392
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-924-035A-470

Query Match      1.0%; Score 18; DB 10; Length 392;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 517 ATTTTATTGTCCTCAAG 534
      |||||||
Db 66 ATTTTATTGTCCTCAAG 83

RESULT 13
US-09-983-965-2819/c
; Sequence 2819, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 2819
; LENGTH: 421
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 15-LIB3058-038-Q1-K1-D3
US-09-983-965-2819

Query Match      1.0%; Score 18; DB 10; Length 421;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 713 ACTAAATTTTCTTCATG 730
      |||||||
Db 268 ACTAAATTTTCTTCATG 251

RESULT 14
US-09-920-300A-17/c
; Sequence 17, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 459
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-17

Query Match      1.0%; Score 18; DB 10; Length 459;
```

```

Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 887 TCCGTTTGGGAATGGAT 904
      |||||||
Db 36 TCCGTTTGGGAATGGAT 19

RESULT 15
US-10-033-528-17/c
; Sequence 17, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033,528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 459
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-17

Query Match      1.0%; Score 18; DB 12; Length 459;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 887 TCCGTTTGGGAATGGAT 904
      |||||||
Db 36 TCCGTTTGGGAATGGAT 19

RESULT 16
US-09-770-444-3/c
; Sequence 3, Application US/09770444
; Patent No. US20020023280A1
; GENERAL INFORMATION:
; APPLICANT: Goxlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krieker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2027 (PARA-016PRV)
; CURRENT APPLICATION NUMBER: US/09/770,444
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,502
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
```

LENGTH: 478  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-770-444-3

Query Match 1.0%; Score 18; DB 10; Length 478;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 835 TCGAGTGTCTTTCTTT 852  
|||||  
DB 97 TCGAGTGTCTTTCTTT 80

RESULT 17  
US-09-770-149-852  
Sequence 852, Application US/09770149  
Patent No. US20020059663A1

GENERAL INFORMATION:  
APPLICANT: Goriach, Jorn  
APPLICANT: An, Yong-Qiang  
APPLICANT: Hamilton, Carol M.  
APPLICANT: Price, Jennifer L.  
APPLICANT: Raines, Tracy M.  
APPLICANT: Yu, Yang  
APPLICANT: Rameaka, Joshua G.  
APPLICANT: Page, Amy  
APPLICANT: Matthew, Abraham V.  
APPLICANT: Ledford, Brooke L.  
APPLICANT: Moesener, Jeffrey P.  
APPLICANT: Haas, William David  
APPLICANT: Garcia, Carlos A.  
APPLICANT: Kicker, Maya  
APPLICANT: Slader, Ted  
APPLICANT: Davis, Keith R.  
APPLICANT: Allen, Keith  
APPLICANT: Hoffman, Neil  
APPLICANT: Hurlan, Patrick  
TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
FILE REFERENCE: 2024 (PARA-013PRV)  
CURRENT APPLICATION NUMBER: US/09/770,149  
CURRENT FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: 60/178,506  
PRIOR FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 999  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 852  
LENGTH: 613  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-770-149-852

Query Match 1.0%; Score 18; DB 10; Length 613;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 517 ATTTTATGTCTCAAG 534  
|||||  
DB 73 ATTTTATGTCTCAAG 90

RESULT 18  
US-09-764-847-1360  
Sequence 1360, Application US/09764847  
Patent No. US20020132767A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC009  
CURRENT APPLICATION NUMBER: US/09/764,847  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 2003  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1360  
LENGTH: 713  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-847-1360

Query Match 1.0%; Score 18; DB 10; Length 713;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 617 GCACCTCTTTCTTCTTA 634  
|||||  
DB 254 GCACCTCTTTCTTCTTA 271

RESULT 19  
US-09-938-842A-5222  
Sequence 5222, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
FILE REFERENCE: SCRIPI300-3  
CURRENT APPLICATION NUMBER: US/09/938,842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 5222  
LENGTH: 755  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-5222

Query Match 1.0%; Score 18; DB 9; Length 755;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1707 TTGCACCACTCTGTGT 1724  
|||||  
DB 367 TTGCACCACTCTGTGT 384

RESULT 20  
US-09-887-576-715  
Sequence 715, Application US/09887576  
Patent No. US20020144047A1  
GENERAL INFORMATION:  
APPLICANT: Budworth, P.  
APPLICANT: Brown, D.  
APPLICANT: Chang, H.  
APPLICANT: Zhu, T.  
APPLICANT: Han, B.  
APPLICANT: Wang, X.  
APPLICANT: Cooper, Bret  
TITLE OF INVENTION: Promoters for regulation of plant expression  
FILE REFERENCE: 1360.001US1  
CURRENT APPLICATION NUMBER: US/09/887,576  
CURRENT FILING DATE: 2001-06-25  
PRIOR APPLICATION NUMBER: US 60/213,848  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: US 60/214,087  
PRIOR FILING DATE: 2000-06-23



APPLICANT: Chen, Wensheng  
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 FILE REFERENCE: Aecmlca-X-1  
 CURRENT APPLICATION NUMBER: US/09/864,761  
 CURRENT FILING DATE: 2001-05-23  
 PRIOR APPLICATION NUMBER: US 60/180,312  
 PRIOR FILING DATE: 2000-02-04  
 PRIOR APPLICATION NUMBER: US 60/207,456  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: US 09/632,366  
 PRIOR FILING DATE: 2000-08-03  
 PRIOR APPLICATION NUMBER: GB 24263.6  
 PRIOR FILING DATE: 2000-10-04  
 PRIOR APPLICATION NUMBER: US 60/236,359  
 PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: PCT/US01/00666  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00667  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00664  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00669  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00665  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00668  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00663  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00662  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00661  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00670  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: US 60/234,687  
 PRIOR FILING DATE: 2000-09-21  
 PRIOR APPLICATION NUMBER: US 09/608,408  
 PRIOR FILING DATE: 2000-06-30  
 PRIOR APPLICATION NUMBER: US 09/774,203  
 PRIOR FILING DATE: 2001-01-29  
 NUMBER OF SEQ ID NOS: 49117  
 SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
 SEQ ID NO 26319  
 LENGTH: 178  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 OTHER INFORMATION: MAP TO AC021498.4  
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5  
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2  
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4  
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1  
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3  
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2  
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9  
 OTHER INFORMATION: EST HUMAN HIT: AL041184.2, EVALUATE 9.60e+00  
 OTHER INFORMATION: SWISSPROT HIT: OS9295, EVALUATE 3.90e+00  
 OTHER INFORMATION: NT HIT: X52046.1, EVALUATE 1.60e+00  
 US-09-864-761-26319

Query Match 0.9%; Score 17; DB 10; Length 178;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 658 TGCTTGAATGATTT 674  
 |||||  
 Db 98 TGCTTGAATGATTT 114

RESULT 26  
 US-10-046-935-1387

Sequence 1387, Application US/10046935  
 Patent No. US2002015601A1  
 GENERAL INFORMATION:  
 APPLICANT: Jiang, Yugu  
 APPLICANT: Harlocker, Susan L.  
 APPLICANT: Secrist, Heather  
 APPLICANT: Wang, Aijun  
 APPLICANT: Stolk, John A.  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 FILE REFERENCE: 210121.527C1  
 CURRENT APPLICATION NUMBER: US/10/046,935  
 CURRENT FILING DATE: 2002-01-15  
 NUMBER OF SEQ ID NOS: 2239  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 1387  
 LENGTH: 293  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-10-046-935-1387

Query Match 0.9%; Score 17; DB 9; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1609 TTGTTAATGCTTTTGT 1625  
 |||||  
 Db 148 TTGTTAATGCTTTTGT 164

RESULT 27  
 US-09-878-178-1387  
 Sequence 1387, Application US/09878178  
 Patent No. US20020177552A1  
 GENERAL INFORMATION:  
 APPLICANT: Jiang, Yugu  
 APPLICANT: Harlocker, Susan L.  
 APPLICANT: Secrist, Heather  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 FILE REFERENCE: 210121.527  
 CURRENT APPLICATION NUMBER: US/09/878,178  
 CURRENT FILING DATE: 2001-06-08  
 NUMBER OF SEQ ID NOS: 2237  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 1387  
 LENGTH: 293  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-878-178-1387

Query Match 0.9%; Score 17; DB 9; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1609 TTGTTAATGCTTTTGT 1625  
 |||||  
 Db 148 TTGTTAATGCTTTTGT 164

RESULT 28  
 US-09-962-436-452  
 Sequence 452, Application US/09962436  
 Patent No. US20020081301A1  
 GENERAL INFORMATION:  
 APPLICANT: Soppet, Daniel  
 TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signa  
 FILE REFERENCE: 689290-75  
 CURRENT APPLICATION NUMBER: US/09/962,436  
 CURRENT FILING DATE: 2001-09-25  
 PRIOR APPLICATION NUMBER: US/60/235,082  
 PRIOR FILING DATE: 2000-09-25



```
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 452
; LENGTH: 324
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-436-452

Query Match          0.9%; Score 17; DB 10; Length 324;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 TCAGATTAAATCTGG 207
   |||||||
DB 153 TCAGATTAAATCTGG 169

RESULT 29
US-09-867-701-6153
; Sequence 6153, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6153
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-6153

Query Match          0.9%; Score 17; DB 10; Length 403;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1783 TGTAGCAATATAAAAT 1799
   |||||||
DB 67 TGTAGCAATATAAAAT 83

RESULT 30
US-09-983-965-3885
; Sequence 3885, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 3885
; LENGTH: 434
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
```

```
; OTHER INFORMATION: Clone ID: 50-LIB3058-007-Q1-K1-E6
US-09-983-965-3885

Query Match          0.9%; Score 17; DB 10; Length 434;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 843 TTTTCTTTTAGGTGC 859
   |||||||
DB 366 TTTTCTTTTAGGTGC 382

RESULT 31
US-09-867-701-5590/c
; Sequence 5590, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5590
; LENGTH: 478
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-5590

Query Match          0.9%; Score 17; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1459 GTATTCAATAACAATA 1475
   |||||||
DB 423 GTATTCAATAACAATA 407

RESULT 32
US-09-998-598-1850/c
; Sequence 1850, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Meagher, Madelein Joy
; APPLICANT: Chenault, Ruth A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 1850
; LENGTH: 478
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-1850

Query Match          0.9%; Score 17; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 426 CACTCACCTTGTGAGGA 442
   |||||||
DB 433 CACTCACCTTGTGAGGA 417
```

RESULT 33  
US-09-864-761-10630/c  
Sequence 10630, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecm1ca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 10630  
LENGTH: 494  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC005392.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.91  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.9  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.94  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.77  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.99  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.8  
US-09-864-761-10630

Query Match 0.9%; Score 17; DB 10; Length 494;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1319 TGGCAATATGCTTCAG 1335  
|||||  
Db 263 TGGCAATATGCTTCAG 247

RESULT 34  
US-09-734-569-67  
Sequence 67, Application US/09734569  
Patent No. US20020064816A1  
GENERAL INFORMATION:  
APPLICANT: Lerchl, Jens  
APPLICANT: Renz, Andreas  
APPLICANT: Ehrhardt, Thomas  
APPLICANT: Reindel, Andreas  
APPLICANT: Cippus, Petra  
APPLICANT: Bischoff, Friedrich  
APPLICANT: Freund, Markus  
APPLICANT: Duwendt, Elke  
APPLICANT: Schmidt, Ralf-Michael  
APPLICANT: Reski, Ralf  
TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involve  
FILE REFERENCE: BASF-NAE-1332-99-US  
CURRENT APPLICATION NUMBER: US/09/734,569  
CURRENT FILING DATE: 2001-05-24  
PRIOR APPLICATION NUMBER: US 60/171,101  
PRIOR FILING DATE: 1999-12-16  
NUMBER OF SEQ ID NOS: 181  
SOFTWARE: PatentIn Ver. 2.1/WordPerfect 6.1  
SEQ ID NO 67  
LENGTH: 496  
TYPE: DNA  
ORGANISM: Physcomitrella patens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(297)  
OTHER INFORMATION: 94\_ck14\_hilfwd  
US-09-734-569-67

Query Match 0.9%; Score 17; DB 10; Length 496;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1572 GCCGACGCTAATGACA 1588  
|||||  
Db 146 GCCGACGCTAATGACA 162

RESULT 35  
US-09-864-761-7444  
Sequence 7444, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecm1ca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117

```

```

; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7444

```

```

; SEQ ID NO 7444
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Homo sapiens
;
; FEATURE:
;
;   OTHER INFORMATION: MAP TO AC021498.4
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; US-09-864-761-7444

```

Query Match	0.9%	Score 17;	DB 10;	Length 539;
Best Local Similarity	100.0%;	Pred. No. 71;		
Matches 17;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 658 TGCTTTGACATTGATT 674  
|||  
Db 173 TGCTTTGACATTGATT 189

```

RESULT 36
US-09-734-569-65
US-09-734-569-65, Application US/09734569
; Patent No. US20020064816A1
; GENERAL INFORMATION:
; APPLICANT: Lerchl, Jens
; APPLICANT: Renz, Andreas
; APPLICANT: Erhardt, Thomas
; APPLICANT: Reindl, Andreas
; APPLICANT: Cirpus, Petra
; APPLICANT: Bischoff, Friedrich
; APPLICANT: Frank, Markus
; APPLICANT: Freund, Annette
; APPLICANT: Duwenig, Elke
; APPLICANT: Schmidt, Ralf-Michael
; APPLICANT: Reski, Ralf
; TITLE OF INVENTION: Moss genes from
; FILE REFERENCE: BASF-NAE-1332-99-US
; CURRENT APPLICATION NUMBER: US/09/7
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 60/171
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 181

```

```

; SOFTWARE: PatentIn Ver. 2.1/WordPerfect 6.1
; SEQ ID NO 65
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Physcomitrella patens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(375)
; OTHER INFORMATION: 71_ck18_d06fwd
US-09-734-569-65

```

Query Match 0.9%; Score 17; DB 10; Length 548;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 17; Conservative 0; Mismatches 0; Indels

QY 1572 GCCGCACGGTAATGACA 1588  
|||  
Db 224 GCCGCACGGTAATGACA 240

```

RESULT 37
US-09-969-347-157
; Sequence 157, Application US/09969347
; Patent No. US20020115085A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Cancer Gene Determina
; TIME OF INVENTION: Sets
; FILE REFERENCE: 689290-69
; CURRENT APPLICATION NUMBER: US/09/969,347
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/60/237,598
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,604
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 318
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 157
; LENGTH: 599
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: n=a,t,g or c
US-09-969-347-157

```

Query Match 0.9%; Score 17; DB 10; Length 599;  
Best Local Similarity 100.0%; Pred.No. 71;  
Matches 17; Conservative 0; Mismatches 0; Indels

Qy 191 TCAGATTAAATCTGG 207  
Db 157 TCAGATTAAATCTGG 173

```

RESULT 38
US-09-999-162-9
; Sequence 9, Application US/09999162
; Patent No. US20020137171A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Mathur, Eric J.
; APPLICANT: Baumann, Markus
; APPLICANT: Bornscheuer, Uwe T.
; TITLE OF INVENTION: Hydrolase Enzymes and Their Use in Kinetic Resolution
; FILE REFERENCE: DIV-011 US/112766.126
; CURRENT APPLICATION NUMBER: US/09/999,162
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: US 60/250,976
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9

```

LENGTH: 933  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: polynucleotide BD073 encoding polypeptide with  
OTHER INFORMATION: hydrolase activity  
US-09-999-162-9

Query Match 0.9%; Score 17; DB 10; Length 933;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 17; Conservative 0; Indels 0; Gaps 0;

QY 1232 CTATCGAAGAAATGCG 1248  
DB 587 CTATCGAAGAAATGCG 603

RESULT 39  
US-09-734-569-161  
Sequence 161, Application US/09734569  
Patent No. US20020064816A1  
GENERAL INFORMATION:

APPLICANT: Lerchl, Jens  
APPLICANT: Renz, Andreas  
APPLICANT: Ehrhardt, Thomas  
APPLICANT: Reindl, Andreas  
APPLICANT: Cirus, Petra  
APPLICANT: Bischoff, Friedrich  
APPLICANT: Frank, Markus  
APPLICANT: Freund, Annette  
APPLICANT: Duwenig, Elise  
APPLICANT: Schmidt, Ralf-Michael  
APPLICANT: Reski, Ralf  
TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved  
TITLE OF INVENTION: in the synthesis of carbohydrates  
FILE REFERENCE: BASF-NAE-1332-99-US  
CURRENT APPLICATION NUMBER: US/09/734,569  
CURRENT FILING DATE: 2001-05-24  
PRIOR APPLICATION NUMBER: US 60/171,101  
PRIOR FILING DATE: 1999-12-16  
NUMBER OF SEQ ID NOS: 181  
SOFTWARE: Patentin Ver. 2.1/WordPerfect 6.1  
SEQ ID NO 161  
LENGTH: 979  
TYPE: DNA  
ORGANISM: Physcomitrella patens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(798)  
OTHER INFORMATION: c\_dp001074086r  
US-09-734-569-161

Query Match 0.9%; Score 17; DB 10; Length 979;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1572 GCCGCACGATATGACA 1588  
DB 647 GCCGCACGATATGACA 663

RESULT 40  
US-09-938-842A-5070  
Sequence 5070, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:

APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
TITLE OF INVENTION: SAME, AND METHODS OF USE  
FILE REFERENCE: SRIPI300-3

CURRENT APPLICATION NUMBER: US/09/938,842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 5070  
LENGTH: 1121  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-5070

Query Match 0.9%; Score 17; DB 9; Length 1121;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1228 CAAACTATCGAAGAAA 1244  
DB 988 CAAACTATCGAAGAAA 1004

RESULT 41  
US-09-815-242-9546/c  
Sequence 9546, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zykand, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Tiawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9546  
LENGTH: 1551  
TYPE: DNA  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1551)  
US-09-815-242-9546

Query Match 0.9%; Score 17; DB 10; Length 1551;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1762 ACCGTAACCTATATCT 1778  
|||||

Db 1239 ACCGTTAAACTTAATCT 1223

RESULT 42

US-09-939-980-116/c  
; Sequence 116, Application US/09939980  
; Patent No. US2002008234A1

GENERAL INFORMATION:

APPLICANT: Black, Michael  
Burnham, Martin  
Hodgson, John  
Knowles, David  
Lonetto, Michael  
Nicholas, Richard  
Pratt, Julie  
Reichard, Richard  
Rosenberg, Martin  
Ward, Judith

TITLE OF INVENTION: No. US2002008234A1el Prokaryotic Polynucleotides,  
Polypeptides and Their Uses

NUMBER OF SEQUENCES: 534

CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/939,980  
FILING DATE: 27-Aug-2001

PRIOR APPLICATION DATA:

CLASSIFICATION: <Unknown>  
APPLICATION NUMBER: 08/936,165  
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Gimmi, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 116:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1789 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 116:

US-09-939-980-116

Query Match 0.9%; Score 17; DB 10; Length 1789;

Best Local Similarity 100.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 CCTCAGATTTAAATCT 205

|||||

Db 94 CCTCAGATTTAAATCT 78

RESULT 43

US-09-887-576-173  
; Sequence 173, Application US/09887576  
; Patent No. US20020144047A1

GENERAL INFORMATION:

APPLICANT: Budworth, P.

APPLICANT: Brown, D.  
APPLICANT: Chang, H.  
APPLICANT: Zhu, T.  
APPLICANT: Han, B.  
APPLICANT: Wang, X.  
APPLICANT: Cooper, Bret

TITLE OF INVENTION: Promoters for regulation of plant expression  
FILE REFERENCE: 1360.001US1

CURRENT APPLICATION NUMBER: US/09/887,576

CURRENT FILING DATE: 2001-06-25

PRIOR APPLICATION NUMBER: US 60/213,848

PRIOR FILING DATE: 2000-06-23

PRIOR APPLICATION NUMBER: US 60/214,087

PRIOR FILING DATE: 2000-06-23

PRIOR APPLICATION NUMBER: US 60/258,692

PRIOR FILING DATE: 2000-12-29

NUMBER OF SEQ ID NOS: 875

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 173

LENGTH: 2004

TYPE: DNA

ORGANISM: Arabidopsis thaliana

US-09-887-576-173

Query Match 0.9%; Score 17; DB 10; Length 2004;

Best Local Similarity 100.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 618 CACTTCTTTCTTCCTA 634

|||||

Db 110 CACTTCTTTCTTCCTA 126

RESULT 44

US-09-950-788-1/c

; Sequence 1, Application US/09950788

; Patent No. US20020132323A1

GENERAL INFORMATION:

APPLICANT: MOECKEL, Bettina

APPLICANT: HANS, Stephan

APPLICANT: SCHISCHKA, Natalie

APPLICANT: PFEFFERLE, Walter

TITLE OF INVENTION: NUCLEOTIDE SEQUENCES WHICH CODE FOR THE PTSI PROTEIN

FILE REFERENCE: 213544US0

CURRENT APPLICATION NUMBER: US/09/950,788

CURRENT FILING DATE: 2001-09-13

PRIOR APPLICATION NUMBER: DE10045496.8

PRIOR FILING DATE: 2000-09-13

NUMBER OF SEQ ID NOS: 9

SOFTWARE: PatentIn version 3.1

SEQ ID NO 1

LENGTH: 2005

TYPE: DNA

ORGANISM: Corynebacterium glutamicum

FEATURE:

NAME/KEY: CDS

LOCATION: (154)..(1857)

OTHER INFORMATION:

US-09-950-788-1

Query Match 0.9%; Score 17; DB 10; Length 2005;

Best Local Similarity 100.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1027 TATTCAAACATTTGGAT 1043

|||||

Db 91 TATTCAAACATTTGGAT 75

RESULT 45

US-09-950-788-3/c

; Sequence 3, Application US/09950788

; Patent No. US20020132323A1

```
; GENERAL INFORMATION:
; APPLICANT: MOECKEL, Bettina
; APPLICANT: HANS, Stephan
; APPLICANT: SCHISCHKA, Natalie
; APPLICANT: PEEFERLE, Walter
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES WHICH CODE FOR THE PTSI PROTEIN
; FILE REFERENCE: 213544US0
; CURRENT APPLICATION NUMBER: US/09/950,788
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: DE10045496.8
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2005
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (154)..(1857)
; OTHER INFORMATION:
US-09-950-788-3

Query Match          0.9%; Score 17; DB 10; Length 2005;
Best Local Similarity 100.0%; Pred.No.80;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1027 TATTCAACACATTGGAT 1043
   |||||
Db  91 TATTCAACACATTGGAT 75
```

Search completed: December 17, 2002, 11:36:00  
Job time : 148 secs

```

MEDLINE 99206606
PUBMED 10192398
REFERENCE 2 (bases 1 to 14134)
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L.,
Grimwood,J., Davis,R.W. and Stephens,R.S.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
source
Location/Qualifiers
1. .14134
/organism="Chlamydia pneumoniae CWL029"
/strain="CWL029"
/db_xref="taxon:115713"
complement (158..1924)
/gene="recJ"
/note="CpN0563"
complement (158..1924)
/gene="recJ"
/codon_start=1
/transl_table=11
/product="ssDNA Exonuclease"
/protein_id="AAD18703.1"
/db_xref="GI:4376856"
/translation="MTNSDNAAGLLWAHKEPDAFLGMIKEFHLPTVAQIFIGR
GFOTIOETHKLYSHLSLVDPLGFLDMKAVERLLARDRKEHVMYIGSDVDGWTG
VALLVEFLRDIDVHVSYPFLGAILKHQGETSTLAKLKEGITILLITVDGIGTAGKEV
SVTRQGDIVDITDHHMETGKIPHCVTAINPKRDHTYFNRELIGVGVAFKARGVLN
ALISRNLPKDSQGSLLKLDLVTGTTDVGVLGGENRMVYVYGIKEIKARGLANK
LCAICGVSESVTSDIVLKIAPKLSLGRDDPAKGVLELLTQDDERVDALIMELDN
INRQRTEAEVFDVQVILNSPEILKQAAIILVLSSTAMHARVPIIISARLAKTYNKP
VVIILAOIGCKGARTIGSPFLGLVKKSSLLSYGCHDFAAAGVIMKEKVDKDEKK
KVFHLVNSLKKGTDLPHLEIDAYADFAIDYDILLASMELEFPFGKGNLMPFIYSKVR
QVRYPLVPLGKHLKYLQSKRNLGEGVAGLGRHADALKAWSHYPLEIATYTPRUSQTS
GSGVHLLVRDRFISSEPRFSD"
complement (2010..6218)
/gene="secD/secF"
/note="CpN0564"
complement (2010..6218)
/gene="secD/secF"
/codon_start=1
/transl_table=11
/product="Protein Export Proteins SecD/SecF (fusion)"
/protein_id="AAD18704.1"
/db_xref="GI:4376857"
/translation="MKQVKKNFAIIICVFALALYYVLPTCLYVAKPLKKIDGNEAB
HIKFTQAOQVRKDVIPRVSAILSSHLRGHQHPAIDIVSVRFRGEDAEDFI
GNLHGEENVPIKARLHVGYSRHDDHVIQVASSINTSLVESDFSVYSSENEOE
MASSILQRYVYACTPKQKDCSCYPSIETAPKQLQYAKNLSGFEVPSRLSAP
COQFSSNODRLAFLRSLSSNDAAIDVEDQKLKSVETLSOTACIRSLDCPYIEG
LRLDSESLSFSSIEYCPKERRKIFLTHSLLAQTQSLSKERLDFPSRLAVKQKL
SKNLTUVQEDYNGFSQWMDKTOGKIILOGERLLOGLIAHLTALTUHRPAASCDL
IPENPVFCRQPRESEAFGCYIFSTDCRFGSKGVSILLKGRISIVAKYQQGGKE
LQSFEDLQNLNFCPSHTEAISWTGLGQVLEIRHPLQQLDVMGEGFVKEGCAFL
EVKIDQDLATVNOIEKRNQSDLVWHEQYRHAKSMOLQRLSAPIYONLFLENK
LNMRKFSRGENILRLIGIDFVGROLLLSFKDHQKQLTDKEDILKVSDELKRLG
VSEILRSGEYIHLVSVPGSTISSEILGTSKSFHVNNRPSYSASRYEVRQLD
LYWFTSQAGKTSVPBEINTFASALNEEDVPPSVHEAITLKLSEGLAFSPSGCETLS
TDLDTFMSIAKDEAQKAPLVIFRNYALDGLASLKDRIPEFAAGGYYLVNFSVD
TSPPKMAKSLDPTESFHTWTSAYCOEGISGTANGQYSANRGMWAVIDGVMVSPIL
NVLKNHASVSGKTHREVSKLASDLKSGAMSFPVEVLSEETISDLGKCKCTGGIS
ACCGIAMLIVMSVYRFGVTSAGVAILNLLIWAALOYLDAPLTLGSLGIVLWG
MAVDANLVERIREPRLLSQKLKSVKGYTKAFCAIFDSNLTVTLASALLPFLDTG
PIKGALLTILIGIFSSMFTALFMTKFFMLWNKQTOQLHMMNKFGVGHDFLRGK
KLWAGSVFLGLGVALGAWNSVLGMDFKGYAFTFNPKHEGIDSVAQMRGVVHK
LQEAGLSKDRFIOQFGSSEKIKYFSDKALSYTKADTSLSPKINDHELAVGLLSE
TGLDFTSTLNTQNTFWSKSKMRYOATIGLIGALAIILLYVSLRFWEQYVAFS
AVCALIHDLATCAVLAFLTAHFLKXIQLDQAGLMTVLGYSLNNTLIIEDRIEOR
QANLTPHVLNVALQKTFSRVTNTATISVLMLLFIQSSSVNFNFIWITGILL
GTLSSLYTAPPLLLFWKRNRSK"
complement (6398..7498)
/gene="CpN0565"
complement (6398..7498)
/gene="CpN0565"

```

---

```

/codon_start=1
/transl_table=11
/product="Ct449 hypothetical protein"
/protein_id="AAD18705.1"
/db_xref="GI:4376858"
/translation="MKFPRISDLIPTOMVIMRGGNVHVPNAQLPKKILGGVL
ACFGLALGCAAPAGVCOTIIPCCIGLMLTGLVGLGFAVLOYSKGMSFEPPLRETK
VPEKPNWLGCLSLQGSWKIRPGCYHYHFGCQVEICEGSEIYTKIFOKKSDRNTSI
FLIQEMDQIALRGKIEKSLSRKTFATDPSVSSVSSLSSEIQREEQQLDLPKVISWSSD
QASDRTHPKSAIYVNLSDAAPQGRCYDAYTKAFFTVLDQIGDPNIVKGTIYVLT
PILGVDPALPKEOEENKLLSQALFSAEQVAKMRREKQDSIRIKETFTDPTSPTS
LYFSPHSSTPHSVTPISLSGFGVEQESYTF"
8009..8761
/gene="yaeS"
/note="CpN0566"
8009..8761
/gene="yaeS"
/codon_start=1
/transl_table=11
/product="yaeS family hypothetical protein"
/protein_id="AAD18706.1"
/db_xref="GI:4376859"
/translation="MSLATNNAESKPPSLQRLPNHVAIIMDGNRRWYKKBRECCGHTH
TSGHYGAQVLPNINLAVLDLGLKVLTVTFSTENFRPKKEIOEINFIYTDLDKQL
PYLMENEICLCIGDLSKLPKGIQTKINHVSMTASFSLLELVAVNGKQELVRAF
KXHLVDILNKKISSDLSLSSGLDTPDLDLIRTGEMRVSNFLMQIAYTE
LYITDULWDPDFTPDQLFEAINVYQKSRGGK"
8762..9688
/gene="cdaA"
/note="CpN0567"
8762..9688
/gene="cdaA"
/codon_start=1
/transl_table=11
/product="Phosphatidate Cytidylyltransferase"
/protein_id="AAD18707.1"
/db_xref="GI:4376860"
/translation="MLNSKFKSKTGAYGDLQFQVVVHSLVTLFVLLVSSLPPLTS
FALGPITACGAVGTVEYSMAKAMKHPLSTFSAIGSFLFALSLSIRWGHSLPFGF
FDALPWTLLVVMVYSIFRKRSTIGALQSGVTLFSLYGVIGPILFLVLYSFIHT
QESYGLIWMASFIATTKGADIFGVFGKAFGNKKIAPQISPNKTVGVFVAGCLGATL
ISPIFLQIPTRFASYPMPAILIPLGLALGITGFGFDIIESIKRDAHLKNSNKLKA
VGMGLDLSLLSTPIAYLFLITQSKFEG"
9685..10335
/gene="cmk"
/note="CpN0568"
9685..10335
/gene="cmk"
/codon_start=1
/transl_table=11
/product="CMP kinase"
/protein_id="AAD18708.1"
/db_xref="GI:4376861"
/translation="MIITIDGSGTGKSTAKALADHLHFNVCNTGMVRYTLAVARLQ
SPWATPLPKLEEPSESPTATQGLSEFPFNCHLLTSELTTQEVANAASELSQLEPV
RAPQQLQRYAQLGNCVPEGRMDGKVPFNADLKIFLTSSPEVRAQRKUDLPEGLT
SPQLQALRYKDAADAQRAHDLPLVPENGVIDSDSLTIRQVLEKILALLFRLNEL"
10332..10970
/gene="plsC"
/note="CpN0569"
10332..10970
/gene="plsC"
/codon_start=1
/transl_table=11
/product="Glycerol-3-P Acyltransferase"
/protein_id="AAD18709.1"
/db_xref="GI:4376862"
/translation="MIFRICKFTTWAFSLFYKLKYGVKNFIKGPALIVNHNHSL
DPIALHMCVHECIYHLARSLFNPWLQWQCFVRQDEGNSAFKIASRLFNKRRK
LVITYPEGARSPDQLOPKGVIGMAAKSRVPIIPVYIRGTPEAFNRHOKIPIHWKTI
TCVFGTMYFDIIQNPFIKNKETYQIITNOTWNKIAELKAWYESCKGDVP"
10975..12660
/gene="argS"

```

**THIS PAGE BLANK (USPTO)**



CDS	/note="Cp00570"		
	10975..12660	/protein_id="AAD18710.1"	
	/gene="argS"	/db_xref="GI:4376863"	
	/codon_start=1		
	/transl_table=11		
	/product="arginyl tRNA Transferase"		
	/protein_id="AAD18710.1"		
	/db_xref="GI:4376863"		
	/translation="MSTLLSILSVICQAIKAPFNLEDWAFETPTSPKHEFGHYQCNDAMKLAVLKKAIPRAIAEIVAEIPOEFSLEIETAGAGFINFTSPVFNLOOHLHFKDALKLGFOVSOPKKIILIDPSSNIAKMDHVGHILRSTIIGDSLARIFSYVGHVDVLRNLHIDQWATAGMLITYIQENPCDYSLELDTLSLYKAYVCFYNDEEPKRSQOQNVLAQAKDPQAGIACMETSEKSFQKIDILDIVBKRSEFAYNPFPEIIEOLEKGLLTVSDNACVCFHEAFSIFPMQSGDGVNYATIDLAAMYRIEEDHADKIIVITDLGOSLHFQLLEBATAAAGYLOIPGFSGHVGFLDPQQKLRSCGVNKLRELLDTAIEKAEELAREHRPELTDRAIPERAVIGNAIKSYDLSHRSTSYVFPFKMLREGNTAMFLLYAYRLQGIKRSLGFSQLSEGPPEIQEPAEELLALTILRPEALESTIKELCPHPLTDLYNLTHKFNQFFRDSHIDQSPYAKSRILFLCALAEQVILATGHHLLGLTLERL"		
gene	complement (12614..13990) /gene="murA" /note="Cp00571"		
CDS	complement (12614..13990) /gene="murA"		
	/codon_start=1		
	/transl_table=11		
	/product="UDP-N-Acetylglucosamine Transferase"		
	/protein_id="AAD18711.1"		
	/db_xref="GI:4376864"		
	/translation="MQIARQFCGRLNGEVKSGAKNAATKLLVASLLSDOKCTLRNVPDIGDVLTEUCKSLGARHVSMDKETEVELEIYTPICOTRVPTFSNNVRIFILLGALHLPGRCEGYVETPQGDGAIERTLNFHFGLKQGVQISSDSGGYAKAPRGLKGNIIDHLPSYGVATENLIIAAITHAKRTVIKNVALEAILEDLVLFQKAGADITDGNKDRIIFGTGSLGVDDHTIIPDKIEAASPGMAVVGSGRVFVRNAKQELIIPFLKMLRSIGGGFLVSESGIVFFOERPLVGVGLEDVHPGFLTDMQOPAVLLSOAGSSVHETVHENRUGYDLHGLQHMGAECQLPHQCLSTKACRYAIGNPHPSAVIHGATPLWASHLVIPLDRAFGPAYMAALIAIEGGSSIIENTHLLDRGYTNWVKLSRUGAKIQIFDMEOEELTTSFKSLALRADSL"		
BASE COUNT	4180 a	2981 c	2805 g 4168 t
ORIGIN			
Query Match	98.1%;	Score 1829.4;	DB 1; Length 14134;
Best Local Similarity	99.8%;	Pred. No. 0;	
Matches 1863; Conservative	0;	Mismatches 1;	Indels 3; Gaps 3;
Qy	1	ATGGACTTCGCGATATTGTTCAGGAGGGGATCAGCGGCACCTGCTAATGACAAATATTCCTGC 60	
Db	3771	ATGGACTTCGCGATATTGTTCAGGAGGGGATCAGCGGCACCTGCTAATGACAAATATTCCTGC 3712	
Qy	61	AAACCGTGATGCGGTATGCTCTAGTGATGACGGTTATATGGTTCAGCAGCCCTATTTT 120	
Db	3711	AAACCGTGATGCGGTATGCTCTAGTGATGACGGTTATATGGTTCAGCAGCCCTATTTT 3652	
Qy	121	AAACGTCCTCATGAAAAATCATGCGCAGTGCTCAGGGAAATTTACCCACGCTGAAGTCAG 180	
Db	3651	AAACGTCCTCATGAAAAATCATGCGCAGTGCTCAGGGAAATTTACCCACGCTGAAGTCAG 3592	
Qy	181	CAAACTCGCCTCAGATTAAAACTCGAGCGATGCTCTTTTGTTCCTCCGAGGTTCTCAGTGA 240	
Db	3591	CAAACTCGCCTCAGATTAAAACTCGAGCGATGCTCTTTTGTTCCTCCGAGGTTCTCAGTGA 3532	
Qy	241	AGAGACGATCTCTTCTGATCTCTGGGAAAAAACAATGTACACAAGSCATTTCTCAGCATG 300	
Db	3531	AGAGACGATCTCTTCTGATCTCTGGGAAAAAACAATGTACACAAGSCATTTCTCAGCATG 3472	
Qy	301	CTGTGGCTTGGCAATGCTTATGTTTTGATGAGCGGTATATTATAGATTGAGGCGGTCAT 360	
Db	3471	CTGTGGCTTGGCAATGCTTATGTTTTGATGAGCGGTATATTATAGATTGAGGCGGTCAT 3412	
Qy	361	CGCTTCGGGAGCTCTTCTTCGATCTTTTGCTTATCTGGGCAGCTCTACAGTATTTTGA 420	
Db	3411	CGCTTCGGGAGCTCTTCTTCGATCTTTTGCTTATCTGGGCAGCTCTACAGTATTTTGA 3352	
Qy	421	TGCGCCACTACACCTTGTGAGGACTCGCTGGGATTTGTTCTTCTGATGGGGAATGGCGGTAGA 480	

3351	Db	 TGCGCCACTCACCTTGTFCAGACCTCGCTGGGATGTCTTGTCTATGGGGATGGCGGTAGA	3292
481	Qy	 TGCAAATGTTCTTGTAATTCGAAAGAAATCCGAGAGGAAATTTTATTGTCTCTCAAAAGTCCTTAA	540
3291	Db	 TGCAAATGTTCTTGTAATTCGAAAGAAATCCGAGAGGAAATTTTATTGTCTCTCAAAAGTCCTTAA	3232
541	Qy	 AAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGAGGCCATTTTGTATCTTAACCTGCAC	600
3231	Db	 AAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGAGGCCATTTTGTATCTTAACCTGCAC	3172
601	Qy	 TACAGTATTGGCCTCAGCACCTCTTTTCTCTCTAGATACAGGGCCCTATTAAAGGGTTTGC	660
3171	Db	 TACAGTATTGGCCTCAGCACCTCTTTTCTCTCTAGATACAGGGCCCTATTAAAGGGTTTGC	3112
661	Qy	 TTTGACATTGATTTTAGGAAATTTTCTCTTCAATGTTTACGGCTCTTTTCATGACTAAATTT	720
3111	Db	 TTTGACATTGATTTTAGGAAATTTTCTCTTCAATGTTTACGGCTCTTTTCATGACTAAATTT	3052
721	Qy	 TTTCTTCATGCTGTGGATGAATAAGACCCAAACATACACAGTTCGCATATGATGAATGAAGTT	780
3051	Db	 TTTCTTCATGCTGTGGATGAATAAGACCCAAACATACACAGTTCGCATATGATGAATGAAGTT	2992
781	Qy	 CGTGGGGATAAGCATGATTTCTTTGAGAGATGCAAAAAATTTTGGCTGTTTCTCGGAAG	840
2991	Db	 CGTGGGGATAAGCATGATTTCTTTGAGAGATGCAAAAAATTTTGGCTGTTTCTCGGAAG	2932
841	Qy	 TGTTTTTCTTTTAGGTTCGGTGTCTCTCGGTTTGGAGCCTGGAAATTCGGTTTGGGAAT	900
2931	Db	 TGTTTTTCTTTTAGGTTCGGTGTCTCTCGGTTTGGAGCCTGGAAATTCGGTTTGGGAAT	2872
901	Qy	 GGATTTTAAAGAGGGGTATGCTTTTAACTCAAAAGAGCATGGCATCAGCCATGT	960
2871	Db	 GGATTTTAAAGAGGGGTATGCTTTTAACTCAAAAGAGCATGGCATCAGCCATGT	2812
961	Qy	 TGCTCAATTCGGTGCRAAGTTGTGCATTAACATACAGGAAGCTGGTCTTTCTCTAGAGA	1020
2811	Db	 TGCTCAATTCGGTGCRAAGTTGTGCATTAACATACAGGAAGCTGGTCTTTCTCTAGAGA	2752
1021	Qy	 CTTCCGTATTCAAAACATTTGGATCTTCAGAAAAAGATCAAAATCTATTTTAGTGATAAAGC	1080
2751	Db	 CTTCCGTATTCAAAACATTTGGATCTTCAGAAAAAGATCAAAATCTATTTTAGTGATAAAGC	2692
1081	Qy	 TTTAAAGCTATCT - AAGCAGATACGAGCCTCTCTCTTAAATTAACGATCATGAGCTGGC	1139
2691	Db	 TTTAAAGCTATCTATAAAGCAGATACGAGCCTCTCTCTTAAATTAACGATCATGAGCTGGC	2632
1140	Qy	 GTTA - TTGTGGGATTTGTGCAGAAACAGGCTAGATTCTCTACGGAAACTCTTAAACGA	1198
2631	Db	 GTTAGCTGTGGGATTTGTGCAGAAACAGGCTAGATTCTCTACGGAAACTCTTAAACGA	2572
1199	Qy	 AAGCAAAA - TTTTGGTCAAAGGTAAAGCAGCAAACTATCGAAGAAAAATGCGTTATCAGGC	1257
2571	Db	 AAGCAAAAATTTTGGTCAAAGGTAAAGCAGCAAACTATCGAAGAAAAATGCGTTATCAGGC	2512
1258	Qy	 GACCATCGGGCTTTTAGGAGCTTTGGCAATCATCTTGCTCTATGTGAGTTTGCCTTTGA	1317
2511	Db	 GACCATCGGGCTTTTAGGAGCTTTGGCAATCATCTTGCTCTATGTGAGTTTGCCTTTGA	2452
1318	Qy	 ATGGCAATATGCTTTTCAGTGCAGTATGGCTTTAAATTCATGACCTTTGGCTACCTGTC	1377
2451	Db	 ATGGCAATATGCTTTTCAGTGCAGTATGGCTTTAAATTCATGACCTTTGGCTACCTGTC	2392
1378	Qy	 AGTCTTGTTTATAGCACATTTCTTTTGAAGAAAAATTCAAAATAGATTTCGAAGCCATTGG	1437
2391	Db	 AGTCTTGTTTATAGCACATTTCTTTTGAAGAAAAATTCAAAATAGATTTCGAAGCCATTGG	2332
1438	Qy	 TGCTTTAAATGACTGTATTGGGGTATTCATTAAACAAATACTTTGATCACTTTTGTATCGTAT	1497
2331	Db	 TGCTTTAAATGACTGTATTGGGGTATTCATTAAACAAATACTTTGATCACTTTTGTATCGTAT	2272
1498	Qy	 TCGTGAAGATTCGCAAGCGAACTGTTTACCCTATGCATGTTTTAGTATGATGCGCT	1557

**THIS PAGE BLANK (USPTO)**

```

Db 2271 TCGTGAAGATCCCAAGCAACCTGTTTACCCCTATGCATGTTTGTAGTTAATGATGCCT 2212
QY 1558 TCAAAAGAGCTTACGCGCACCGTAATGACAAACAGCTCAACTATCAGTTTGTGTTAAT 1617
|||||
Db 2211 TCAAAAGAGCTTACGCGCACCGTAATGACAAACAGCTCAACTATCAGTTTGTGTTAAT 2152
QY 1618 GCTTTGTTTATAGCGGCTCTCTGCTTAAATTTTCATTTATATGACCAATAGGAT 1677
|||||
Db 2151 GCTTTGTTTATAGCGGCTCTCTGCTTAAATTTTCATTTATATGACCAATAGGAT 2092
QY 1678 TCTTCTAGGAACCTTATCGTCTCTTTATATGACCAACCTCTGTTGTTGTTTATGTCGCG 1737
|||||
Db 2091 TCTTCTAGGAACCTTATCGTCTCTTTATATGACCAACCTCTGTTGTTGTTTATGTCGCG 2032
QY 1738 TAAAGAAATCGCTCAAAATAGTACCGTTAACTTAATCTAACCTGTAGCAATATAAAA 1797
|||||
Db 2031 TAAAGAAATCGCTCAAAATAGTACCGTTAACTTAATCTAACCTGTAGCAATATAAAA 1972
QY 1798 ATCTCTCTTTGGGACCTTATAGTCCCAAGGCCCTGTGGTATTAAATTTATGACAAATTCAG 1857
|||||
Db 1971 ATCTCTCTTTGGGACCTTATAGTCCCAAGGCCCTGTGGTATTAAATTTATGACAAATTCAG 1912
QY 1858 ATAATGC 1864
|||||
Db 1911 ATAATGC 1905

RESULT 2
LOCUS AP002547/c 300550 bp DNA linear BCT 25-MAY-2002
DEFINITION Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,
section 3/4.
ACCESSION AP002547 AB033786 AB033787 AB033816 AB033817 BA000008
VERSION AP002547.2 GI:10176693
KEYWORDS Chlamydomophila pneumoniae J138 (strain:J138) DNA.
SOURCE Chlamydomophila pneumoniae J138
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
REFERENCE 1 Shirai,M., Hirakawa,H., Ouchi,K., Tabuchi,M., Kishi,F., Kimoto,M.,
Takeuchi,A., Nishida,J., Shibata,K., Fujinaga,R., Yoneda,H.,
Matsushima,H., Tanaka,C., Furukawa,S., Miura,K., Nakazawa,A.,
Ishii,K., Shiba,T., Hattori,M., Kuhara,S. and Nakazawa,T.
Comparison of outer membrane protein genes omp and pmp in the whole
genome sequences of Chlamydia pneumoniae isolates from Japan and
the United States
J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)
20298986
REFERENCE 2 Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,
Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.
Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA
Nucleic Acids Res. 28 (12), 2311-2314 (2000)
JOURNAL 20330349
MEDLINE 3 (bases 1 to 300550)
AUTHORS Shirai,M.
TITLE Direct Submission
Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University
School of Medicine, Department of Microbiology; 1-1-1
Minamikogushi, Ube, Yamaguchi 755-8505, Japan
(E-mail:mshirai@po.cc.yamaguchi-u.ac.jp, Tel:81-836-22-2227,
Fax:81-836-22-2415)
On or before Sep 15, 2000 this sequence version replaced
gi:6172298, gi:6172300, gi:6172396, gi:6172398, gi:8978889.
LOCATION/Qualifiers
1. 300550
/organism="Chlamydomophila pneumoniae J138"
/strain="J138"
/db_xref="taxon:138677"
/note="synonym:Chlamydia pneumoniae (strain J138)"
complement(21..794)
/gene="dapF"
source
gene

```

**THIS PAGE BLANK (USPTO)**

24: em\_ph:\*  
 25: em\_pl:\*  
 26: em\_ro:\*  
 27: em\_sts:\*  
 28: em\_un:\*  
 29: em\_vi:\*  
 30: em\_htg\_hum:\*  
 31: em\_htg\_inv:\*  
 32: em\_htg\_othr:\*  
 33: em\_htg\_mus:\*  
 34: em\_htg\_pln:\*  
 35: em\_htg\_rod:\*  
 36: em\_htg\_mam:\*  
 37: em\_htg\_vrt:\*  
 38: em\_sy:\*  
 39: em\_hgo\_hum:\*  
 40: em\_hgo\_mus:\*  
 41: em\_hgo\_othr:\*

Pred. No. is the number of results predicted by chance to have a

# ALIGNMENTS

RESULT 1  
 AE001641/c  
 LOCUS 14134 bp DNA linear BCT 01-DEC-2000  
 DEFINITION Chlamydia pneumoniae section 57 of 103 of the complete genome.  
 ACCESSION AE001641 AE001363  
 VERSION AE001641.1 GI:4376855  
 KEYWORDS  
 SOURCE Chlamydophila pneumoniae CWL029.  
 ORGANISM Chlamydophila pneumoniae CWL029  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.  
 REFERENCE 1 (bases 1 to 14134)  
 AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hymn,R.W.,  
 Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.  
 TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
 JOURNAL Nat. Genet. 21 (4), 385-389 (1999)

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 06:18:31 ; Search time 3364 Seconds  
(without alignments)  
16125.930 Million cell updates/sec

Title: US-09-868-987-1  
Perfect score: 1864  
Sequence: 1 atggacttcgcattgtc.....atgacaaattcagataatgc 1864

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.rod.\*

36: em.htg.mam.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	1829.4	98.1	14134	1	AE001641	AE001641 Chlamydia
C 2	1829.4	98.1	300550	1	AP002547	AP002547 Chlamydia
C 3	1797.4	96.4	12198	1	AE002179	AE002179 Chlamydia
C 4	773.6	41.5	9989	1	AE001318	AE001318 Chlamydia
C 5	760.8	40.8	12442	1	AE002341	AE002341 Chlamydia
C 6	161.2	8.6	11527	1	AE010580	AE010580 Fusobacte
C 7	153.6	8.2	119836	1	AE001567	AE001567 Helicobac
C 8	148.2	8.0	282183	1	CJ11168X4	AL139077 Campyloba
C 9	146.4	7.9	15058	1	AE000652	AE000652 Helicobac
C 10	144.8	7.8	301550	1	AP003192	AP003192 Clostridi
C 11	127	6.8	12592	1	AE007729	AE007729 Clostridi
C 12	124	6.7	300600	1	AP005369	AP005369 Thermosyn
C 13	122.4	6.6	1557	1	AY078352	AY078352 Rickettsi
C 14	122	6.5	12504	1	AE001219	AE001219 Treponema
C 15	119.6	6.4	279110	1	RPX03	AJ235272 Rickettsi
C 16	119.4	6.4	301550	1	AP003134	AP003134 Staphyloc
C 17	119.4	6.4	333750	1	AP004827	AP004827 Staphyloc
C 18	119.4	6.4	342600	1	AP003363	AP003363 Staphyloc
C 19	118	6.3	45613	6	AX067447	AX067447 Sequence
C 20	115	6.2	10963	1	AE013703	AE013703 Yersinia
C 21	115	6.2	205050	1	AJ414155	AJ414155 Yersinia
C 22	112.6	6.0	10899	1	AE008643	AE008643 Rickettsi
C 23	110.2	5.9	13990	1	AE001166	AE001166 Borrelia
C 24	109.6	5.9	11540	1	AE009509	AE009509 Brucella
C 25	108.8	5.8	315079	1	MPULM03	AL445565 Mycoplasma
C 26	108.2	5.8	1848	6	AX084084	AX084084 Sequence
C 27	108.2	5.8	1848	6	AX395145	AX395145 Sequence
C 28	108.2	5.8	13818	1	U32710	U32710 Haemophilus
C 29	106.4	5.7	349619	1	AP002996	AP002996 Mesorhizo
C 30	103.6	5.6	349980	6	AX413017	AX413017 Sequence
C 31	103.6	5.6	349980	6	AX417045	AX417045 Sequence
C 32	102.8	5.5	12201	1	AE006057	AE006057 Pasteurel
C 33	102.2	5.5	3000	1	RCU89979	US9979 Rhodobacter
C 34	102.2	5.5	5839	6	AX416793	AX416793 Sequence
C 35	102	5.5	240050	1	AL591979	AL591979 Listeria
C 36	100.6	5.4	7687	1	AE012971	AE012971 Chlorobiu
C 37	100	5.4	14261	1	AE004160	AE004160 Vibrio ch
C 38	100	5.4	14567	1	AE000716	AE000716 Aquifex a
C 39	99.6	5.3	2226	6	AX142455	AX142455 Sequence
C 40	99.6	5.3	3511	1	AF269419	AF269419 Staphyloc
C 41	99.6	5.3	3511	6	AX144739	AX144739 Sequence
C 42	99.6	5.3	3705	1	AF269550	AF269550 Staphyloc
C 43	99.6	5.3	3705	6	AX144870	AX144870 Sequence
C 44	99	5.3	4357	1	AE013081	AE013081 Thermoana
C 45	98.2	5.3	11217	1	AE003876	AE003876 Xylella f

ALIGNMENTS

RESULT 1  
AE001641/c  
LOCUS  
DEFINITION Chlamydia pneumoniae section 57 of 103 of the complete genome.  
ACCESSION AE001641 AE001363  
VERSION AE001641.1 GI:4376855  
KEYWORDS  
SOURCE Chlamydia pneumoniae CWL029  
ORGANISM Chlamydia pneumoniae; Chlamydiaceae; Chlamydia.  
REFERENCE  
1 (bases 1 to 14134)  
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,  
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.  
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)

**MEDLINE**  
 99206606  
**PUBMED**  
 2 (bases 1 to 14134)  
**REFERENCE**  
 Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,  
 AUTHORs  
 Greenwood, J., Davis, R. W. and Stephens, R. S.  
**TITLE**  
 Direct Submission  
 Submitted (01-DEC-1998) Program in Infectious Diseases, University  
 of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA  
**JOURNAL**  
 Location/Qualifiers  
**FEATURES**  
 source  
 1. 14134  
 /organism="Chlamydia pneumoniae CWL029"  
 /strain="CWL029"  
 /db\_xref="taxon:115713"  
 /complement (158. .1924)  
 /gene="recJ"  
 /note="Cp0563"  
 /complement (158. .1924)  
 /gene="recJ"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ssDNA Exonuclease"  
 /protein\_id="AAD18703.1"  
 /db\_xref="GI:4376856"  
 /translation="MTNSDNASAAAGLMAHKEDPAFLGMIKEPHLPPTVAQIFSR  
 GPTQIOHKFLYSHSLYDPCGLFDMGKAVSRILARKEHMIYGDSDGMCQ  
 VALVEFLRIDIVHVSFYFLGALILKQGETSTILAKKEGITLLITVDCGTTAGKEV  
 SDITRQIGDIVITDHNMPGKIIPHCVATLNPKLKDHTYPERELTGVGVARGLRGVNLK  
 ALISRLNVPSKQSLKKLLDLVLTGTTDGVLTGENRVVVRGKIEIARFAGLAVLNK  
 LCAICGVESEVTSDIVLKIAPKLNSGRLLDPAKGVELLITQDERVALIMEDN  
 INNERORIEAEVODVOEILNSNPEILKOAATYLSSTAMHARVPIISALATYNNK  
 VVIATIORGIGKGSARTIGSFPLIGYTKRSSLSLISGCHDPAAGVIMKEDKEDKK  
 KFAVLVNSLSKGGDTLPHEIDVADDAIDYDLASMELEFPFGKGNLMPITYSKYR  
 QVAPYKVLPGNHLKLYLSQKERNLEGAFLGRRADALKASWHYPLEIATYPLPSQTS  
 GSGCVHLVDRDFIISSEPRFSD"  
 /complement (2010. .6218)  
 /gene="secD/secF"  
 /note="Cp0564"  
 /complement (2010. .6218)  
 /gene="secD/secF"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Protein Export Proteins SecD/SecF (fusion)"  
 /protein\_id="AAD18704.1"  
 /db\_xref="GI:4376857"  
 /translation="MKQVKRNFPAIIICVPAALAYVLPCLYYAKPLDKKIDGNEAE  
 HIISFTKQAQOVKQVIVPRVSAIILSLHRLGHIOQHPAIPDIVSVRFKGBDAEPI  
 GNLVHGEPNVPISARLHVGVYGRHDDHVIQVASSINTSVESDFSVASENEED  
 MASSIIOVRSACTPPKQDCSCSPSIWETAPKEOLLQAKNLSGSEFVSSRSASF  
 COQSPSNQRLAFLRSLSSLDNAADIVEDQKLKSVYETLSQIACIRSLDCPYIEG  
 LRUDCSSESLPFSISIEYCPKEREKIFLTLHSDLLAQRISLSKEQRLDPSRLAYEQKL  
 SKNLTVQVEDYNNGFSPQMDKDTQKILIQGERLLQGLAHLJALTALHRAESCDL  
 I PENPVPFCRQPRESEAFQCYIFSPNDCGHGSGVYALILKGLRSIYAAVYQGGGGE  
 LOSFEKDLQNLVNCFSHTETAIISWTLGEDQVLEIRHPLQCLDVWGEFVIGKSGKGL  
 EVDKIDORLATVNOIEKRNOSDLVRMHEOYRAHKCSMDLERISAPIPYONLPLENNK  
 LMKRPSRGENIIRLIGIDPYGROLLISFKDHQKCOLTJEDKIDILYSDEICARLNLKG  
 VSEIEKREGEVYIHLVSPGSSITSSSETLTSKSRHVNERSSYSASCTYQRELD  
 YLWFTSOAQKTSPEERINTPASAFLNEEDVPPSVHVAITKLSEGLAFSPSCETPS  
 TDLDTTFSMAIGKDAEOKANPLIVIRYANALQKQIRPEFAAGEGVILNFWSPIL  
 TSPKRAEKLSPTESFHTWTSAVCOEGISGTANGQYANRGMMAVVIDGMVSSPIL  
 NVPLKHNASVSGKTRHEVSLASDLKSLMPLLEVLALYLAALQYLDAPLTLSGLAQLVLMG  
 ACCGLAMLIVMSYTRFGVYASGAVLMLLIMWALQYLDAPLTLSGLAQLVLMG  
 MAVDANLVVERIREEPFLLSQSLKSEVKEYTAFAGIIPDSNLITVLSALFLPLDTG  
 PIKGFALTLLIGIFSSMFTALFMTKFFMLMMNKTOHTOLAMNKFGVIGDLELKGK  
 KLMAVSGSVFLGCGVALGFGAMNSVLMGDFKGVAFTFNKEHGISDVAOMRKVAK  
 LOEGASSRDFRIQTFGSSERKITYFSDKALSTYKADTSLSPKIDHEILALAGLSE  
 TGLDSEPTENETNEMFWSKVSLSKSKMRQATYTGILGALATILLYVSLRPFMOVYS  
 AVCALLHDLATCAVLEFIARFLKLTQIDIGALGALMTVYGLSANTLITPDIRER  
 QANLFTPMHVLVNDALQKTSRTVTMTATYLLSVLMLLFTGSSSVENFAITMTIGILL  
 GTLSLIYAPVLLLFVYRKENRSK"  
 /gene="Cp0565" (7498)  
 /complement (6398. .7498)  
 /gene="Cp0565" (7498)  
 /complement (6398. .7498)  
 /gene="Cp0565" (7498)

/codon\_start=1  
 /transl\_table=11  
 /product="Ct49 hypothetical protein"  
 /protein\_id="AAD18705.1"  
 /db\_xref="GI:4376858"  
 /translation="MKPPIISIDLIPTQMVIMWRGGNVHYVPNAONLKKILGVL  
 ACRGALIGCAAFAGVCOITPPCIIGMILIGVILGPAVLQSKMSRERRLPFBETK  
 VPEKPINMGCSLSLQSNKRTIRGCYTHPGCCQVYICSSQSEIYTKIRPKSDRNTSI  
 FLIQEDQIALQGIKESLSKRTAIDPSVSSLSLSEIQREEQYLDPKVITSMESD  
 QASDRTHPKSAIYVNI SDAAQEPQRCYIDATYKAFVVLDDIGDNPVYKHTIYVL  
 PLIGVDALPKKEQENLKLISQAAFLYSAEYQAKRREKQDSIRIKIFITDPTSPTS  
 LYFSPHSSTPSVTPISLGSFVGESESTFA"  
 8009. .8761  
 /gene="yaeS"  
 /note="Cp0566"  
 8009. .8761  
 /gene="yaeS"  
 /codon\_start=1  
 /transl\_table=11  
 /product="yaeS family hypothetical protein"  
 /protein\_id="AAD18706.1"  
 /db\_xref="GI:4376859"  
 /translation="MSLATNNASKEPSPQLRPNHVAIIMDGNRRMYKKRRECGHTH  
 TSGHYGKAVFLNINAVLDLGIKVLTITFTSTENGRKKEIOELFNIFYTOLDKOL  
 PYLMENETLCRGIDLSLKPGLIOTKINNVSRMTAFSFLSELVAVNNGKDEILRAF  
 KKLHVDILNKRTISSDLSLSLYLDLNSGLDPPDLLIRTGSEKMSVSNLMLQIAYTE  
 LITITDLMFDFLPQDLFAINVTYQRRSGK"  
 8762. .9688  
 /gene="cdaA"  
 /note="Cp0567"  
 8762. .9688  
 /gene="cdaA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phosphatidate Cytidylyltransferase"  
 /protein\_id="AAD18707.1"  
 /db\_xref="GI:4376860"  
 /translation="MLNSKRFKSGTGAAGDLPQVNVHSLVTPVILLVSSLPETS  
 PALGFTLTCGAVTIEYSKMAKAKMHPPLSTFSLAISGLFLALSFLIRKWSLGEF  
 FDLPMPTLITVWVWSIFPRVRSTIGALQSVTLSEIIVYGIPIRLFLVLSFHT  
 QEPVYLGIMWASFLIATTKGADIEGYFFGFAFGNKRIAPDISNKTIVGVACCLGATL  
 ISGIFPELOPTRFASVFPMPALILPLGALGITGFFGDIIESIFKRDALKNKNKLKA  
 VGMDLTDLSLSTPIAVLFLITQSKFIC"  
 9685. .10335  
 /gene="cmk"  
 /note="Cp0568"  
 9685. .10335  
 /gene="cmk"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Cmp Kinase"  
 /protein\_id="AAD18708.1"  
 /db\_xref="GI:4376861"  
 /translation="MIRITDGGSGTGTAKALADHLHFNVCNTGKMYRTLVARLQ  
 SPWATLPLKREBPSEFTFATGQPLSEFENGHLTSELTOEVANASSELQULPEV  
 RARMDLQKRRVQOLNCVYFEGRDMGSKYFPNADIKITFLTSSBVAQRALKULPBEVL  
 SEBQQAEBVKKRDAQRAHDPVLPENGIYDSSDLTIRVLEKITALFRLNEL"  
 10332. .10970  
 /gene="pisc"  
 /note="Cp0569"  
 10332. .10970  
 /gene="pisc"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Glycerol-3-P Acyltransferase"  
 /protein\_id="AAD18709.1"  
 /db\_xref="GI:4376862"  
 /translation="MIRICKFTVWASLFLKLVYCYGKKNFKIGAPALIVANNHSL  
 DPLALHMCVHCETIHLARSLNITPLMWQWCCFVPRQDEGSAFAFKIASRLFNKKK  
 LVIYEGASPDGOLQOPGKVGIMMAKSRVPIIPVYIRGTAEAFNRHQKIPHWKTI  
 TCVFGTPMYFDIIIONPEIKNKETQIITNQTMNKIABELKAWYESGCKGDVP"  
 10975. .12660  
 /gene="args"



```

/notes="Cp00570"
10975..12660
/genes="argS"
/codon_start=1
/transl_table=11
/product="Arginyl tRNA Transferase"
/protein_id="AADI8710.1"
/db_xref="GI:4376863"
/translation="MSTLLLSILSVICQAIKAPPNLEWAPETITSTKBHFGHYQCN
DAMKLARVLKKAAPRAIAEAIIVAEIPQPPFSLIEIAGAGFINFTSPVFLNQOLEHFKD
ALKLGFQVQPKIIIDFSSPNIAKMHVGLHSLRTIIGDSLARIFSVVGHDLVRLNHI
DGMWATGMLTIILOENPCVSDLEDLSLVKKAYVCTNDERFKKRSQONVALQAK
DPAIAIWEKICETSEKAFOKIYDILDIVVEKSGSEFYNFPELLIEDLEKGLLTVS
NDACVPEHAFSPFMVQKSDGGVNYATDOLAMRYIESDHDADKIIVDILGOSLHF
QLLEATAIAGIYQGFISVHVGFLVDPPQKKLKTSGENVKLRLELLDIAEKABEA
LRHREPLTDEALQERAPVGINAIKYSDLSHRTSDYFVSEFKMLRFEGNTAMFLLY
AVYRIQGIKRLSLESGPPEIQEPABEELIALTLRPEALESTIKELCPHFLTLD
YLYNLTHKFNFRDHSIQSPVAKSRFLCALAEQVLATGMHLGLKTLERL"
complement(12614..13990)
/genes="murA"
/notes="murA"
/notes="Cp00571"
complement(12614..13990)
/genes="murA"
/codon_start=1
/transl_table=11
/product="UDP-N-Acetylglucosamine Transferase"
/protein_id="AADI8711.1"
/db_xref="GI:4376864"
/translation="MQIAQVFGCRLNGEVKVSQAKNAATKLLVASLLSDQKCTLRNV
PDIGVSLTVELCKLAHVSWDKETVELIYTPLEIOCTRVPTFESVNVPIRILLGA
LLGRCPGVVPTVGGDAIGERTLNFHFEGLKQLGVQISDSSGGYAKAPRGLKGYI
HLPPSVGATENLILAAHAKGRVTKNVALEAILDLVFLQKAGADITDNDRTID
IFGTGGLGSDHLPIDPKFISAAFCGMAAVVSGRVPVFNKQELLIPFLKRLRSIGGG
FLVSESGIBFQRPRLVGVVLETDVHPGLTDWQOPFAVLLSQOQSSVTHETVHEN
RLGYLHGLHMGAECOLPHOCLSTKACRYAIGNPFPHSVATHGATPLWASHLVIPDLRA
GFAYVMAALIAEGGSGIIEHTLLDRGYTNWVKLSLGAKEIQLIFDMEQELITSPKS
LAURDASL"
BASE COUNT 4180 a 2981 c 2805 g 4168 t
ORIGIN

Query Match 98.14; Score 1829.4; DB 1; Length 14134;
Best Local Similarity 99.88; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

Qy 1 ATGACATTCGCGATATTCGACGAGGGGATCAGCGGCACTGCTAATGACAAATATTCGCG 60
Db 3771 ATGACATTCGCGATATTCGACGAGGGGATCAGCGGCACTGCTAATGACAAATATTCGCG 3712

Qy 61 AAACCGTGGATGGCGTATGGCTGTAGTGAATGACGGTTATATGTCAGCAGCCCTATTTT 120
Db 3711 AAACCGTGGATGGCGTATGGCTGTAGTGAATGACGGTTATATGTCAGCAGCCCTATTTT 3652

Qy 121 AAACCTCCATTTGAAATATGTCAGGCGATGCTCTTGTTCGCGAGGTTCTCAGTGA 240
Db 3651 AAACCTCCATTTGAAATATGTCAGGCGATGCTCTTGTTCGCGAGGTTCTCAGTGA 3592

Qy 181 CAAACTCGCCTCAGATTTAAATCTGAGCGATGCTCTTGTTCGCGAGGTTCTCAGTGA 240
Db 3591 CAAACTCGCCTCAGATTTAAATCTGAGCGATGCTCTTGTTCGCGAGGTTCTCAGTGA 3532

Qy 241 AGAGACGATCTCTTCGTATCTTGGGAAAAAACAATGTACAAAGCATATCTCAGCATG 300
Db 3531 AGAGACGATCTCTTCGTATCTTGGGAAAAAACAATGTACAAAGCATATCTCAGCATG 3472

Qy 301 CTGTGGCTTGGCAATGCTATTTGTTGATGAGCGGTATATATATGATTTGAGGGGTCAT 360
Db 3471 CTGTGGCTTGGCAATGCTATTTGTTGATGAGCGGTATATATATGATTTGAGGGGTCAT 3412

Qy 361 CGCTTCGGGAGCTGTTCTTCTGAATCTTTTGTATCTCTGGGCGAGCTCTACAGTATTTGGA 420
Db 3411 CGCTTCGGGAGCTGTTCTTCTGAATCTTTTGTATCTCTGGGCGAGCTCTACAGTATTTGGA 3352

Qy 421 TGCGCCACTACCTTGTCCAGGACTCGCTGGGATTTGTTCTTCTGCTATGGGATGGCCGTAGA 480

```

```

Db 3351 TGGCCCACTCACCTTGTTCAGGACTCGCTGGGATTTGTTCTTGTCTATGGGATGGCCGTAGA 3292
Qy 481 TGCAAATGTTCTTGTATTTGAAAGAATCCGAGAGGAATTTTTTATTTGTCTCAAGTCTTAA 540
Db 3291 TGCAAATGTTCTTGTATTTGAAAGAATCCGAGAGGAATTTTTTATTTGTCTCAAGTCTTAA 3232
Qy 541 AAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATCTAACTTGAC 600
Db 3231 AAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATCTAACTTGAC 3172
Qy 601 TACAGTATCGSCCTCACACCTCTTTTCTTCTAGATACAGGGCTATTAAGGGTTGC 660
Db 3171 TACAGTATCGSCCTCACACCTCTTTTCTTCTAGATACAGGGCTATTAAGGGTTGC 3112
Qy 661 TTTGACATTTGATTTTAGGAAATTTTCTTCTCAATGTTTACGGCTCTTTTTCATGACTAAAT 720
Db 3111 TTTGACATTTGATTTTAGGAAATTTTCTTCTCAATGTTTACGGCTCTTTTTCATGACTAAAT 3052
Qy 721 TTTCTTCATGCTGTGGATGAATAAGACCCCAACATACACAGTTGCTATGATGAATAAGTT 780
Db 3051 TTTCTTCATGCTGTGGATGAATAAGACCCCAACATACACAGTTGCTATGATGAATAAGTT 2992
Qy 781 CGTGGGATAAAGCATGATTTCTTGAGAGGATGCAAAAAAATTTTGGGCTGTTCTGGAG 840
Db 2991 CGTGGGATAAAGCATGATTTCTTGGAGGATGCAAAAAAATTTTGGGCTGTTCTGGAG 2932
Qy 841 TGTCTTCTTCTTGTAGTTGCGTGTCTCGGTTTGGAGCCTTGGAAATTTCCGTTTGGGAAAT 900
Db 2931 TGTCTTCTTCTTGTAGTTGCGTGTCTCGGTTTGGAGCCTTGGAAATTTCCGTTTGGGAAAT 2872
Qy 901 GGATTTTAAAGAGGGGTATGCTTTTACCTTTAAATCCAAAAAGAGCATGCGCATGTC 960
Db 2871 GGATTTTAAAGAGGGGTATGCTTTTACCTTTAAATCCAAAAAGAGCATGCGCATGTC 2812
Qy 961 TGCTCAATCGGTGGCAAGTGTGCTAACTACAGGAGCTGCTCTTCTTCTTAGAGA 1020
Db 2811 TGCTCAATCGGTGGCAAGTGTGCTAACTACAGGAGCTGCTCTTCTTCTTAGAGA 2752
Qy 1021 CTTCGGTATTTCAAAATTTTGGATCTTCAGAAAAAGATCAAAATCTATTTTGTGATTAAGC 1080
Db 2751 CTTCGGTATTTCAAAATTTTGGATCTTCAGAAAAAGATCAAAATCTATTTTGTGATTAAGC 2692
Qy 1081 TTTAAGCTATACT-AAGCAGATACGAGCCTCTCTCTCTTAAATAATTAACGATCATGAGCTGC 1139
Db 2691 TTTAAGCTATACTAAAGCAGATACGAGCCTCTCTCTCTTAAATAATTAACGATCATGAGCTGC 2632
Qy 1140 GTTA-TTGTGGGATTTGTTCAGAAACAGGCTTAGATTTCTCTACGGAATCTCTAAACGA 1198
Db 2631 GTTAGCTGTGGGATTTGTTCAGAAACAGGCTTAGATTTCTCTACGGAATCTCTAAACGA 2572
Qy 1199 AACGCAAAA-TTTTGGTCAAAGGTAAAGCAAACTATCGAAGAAAAATGCGTTATCAGGC 1257
Db 2571 AACGCAAAAATTTTGGTCAAAGGTAAAGCAAACTATCGAAGAAAAATGCGTTATCAGGC 2512
Qy 1258 GACCATCGGGCTTTAGGAGCTTTGGCAATCATCTGCTCTATGTGAGTTTGGCTTTGA 1317
Db 2511 GACCATCGGGCTTTAGGAGCTTTGGCAATCATCTGCTCTATGTGAGTTTGGCTTTGA 2452
Qy 1318 ATGCAATATGCTTTTCAAGTGGCTATGCGCTTTAAATTCATGACCTTTTGGCTACTGTGC 1377
Db 2451 ATGCAATATGCTTTTCAAGTGGCTATGCGCTTTAAATTCATGACCTTTTGGCTACTGTGC 2392
Qy 1378 AGTCTGTTTATAGCACATTTCTTTTGAAGAAAAATTCAAATAGATTTTGAAGCAATGG 1437
Db 2391 AGTCTGTTTATAGCACATTTCTTTTGAAGAAAAATTCAAATAGATTTTGAAGCAATGG 2332
Qy 1438 TGTCTTAAATGACTGATTTGGGGTATTCATTTAAACAATCTTTTGTATCATTTTGTATCGTAT 1497
Db 2331 TGTCTTAAATGACTGATTTGGGGTATTCATTTAAACAATCTTTTGTATCATTTTGTATCGTAT 2272
Qy 1498 TCGTGAAGATCGCAAGCAACCTGTTTACCCCTATGCTATGCTATGCTTTTGTATGATGCGCT 1557

```





Db 50184 TGTTTTCTTTAGGTTGCGTGTGCTCGGGTTTGAGCCGGAATCCGTTTGGGAA 50125  
 Qy 901 GGAATTTAAAGAGGGATGCTTACCTTTAATCCAAAAGAGATGGCATCAGCGATGT 960  
 Db 50124 GGAATTTAAAGAGGGATGCTTACCTTTAATCCAAAAGAGATGGCATCAGCGATGT 50065  
 Qy 961 TGCTCAAAATCGTGGCAAGTGTGCAATTAACAGAAAGCTGCTCTTTCTTTAAGA 1020  
 Db 50064 TGCTCAAAATCGTGGCAAGTGTGCAATTAACAGAAAGCTGCTCTTTCTTTAAGA 50005  
 Qy 1021 CTTCCTGATTCAAACATTTGGATCTTCAGAAAGATCAAAATCTATTTTATGATTAAGC 1080  
 Db 50004 CTTCCTGATTCAAACATTTGGATCTTCAGAAAGATCAAAATCTATTTTATGATTAAGC 49945  
 Qy 1081 TTTAAGCTATACTAACGAGATACGAGCTCTCTCTTAAATTAAGATCATGAGCTGCG 1139  
 Db 49944 TTTAAGCTATACTAACGAGATACGAGCTCTCTCTTAAATTAAGATCATGAGCTGCG 49885  
 Qy 1140 GTTA-TTGTTGGATGTTGTCAGAAAGAGGCTGATTTCTCTAGGAACTCAAAAGCA 1198  
 Db 49884 GTTAAGCTGTTGGATGTTGTCAGAAAGAGGCTGATTTCTCTAGGAACTCAAAAGCA 49825  
 Qy 1199 AACGCAAAA-TTTTGTCAAAGATAGAGCAAACTATCGAAGAAATGCGTTATCAGGC 1257  
 Db 49824 AACGCAAAAATTTTGTCAAAGATAGAGCAAACTATCGAAGAAATGCGTTATCAGGC 49765  
 Qy 1258 GACCATCGGGCTTTTAAAGAGCTTTGGCAATCATCTTGTCTATGATGCTTGGCTTTGA 1317  
 Db 49764 GACCATCGGGCTTTTAAAGAGCTTTGGCAATCATCTTGTCTATGATGCTTGGCTTTGA 49705  
 Qy 1318 ATGGCAATATGCTTTCAGTGGCGATGCGCTTAAATCATGACCTTTTGGCTACCTGTC 1377  
 Db 49704 ATGGCAATATGCTTTCAGTGGCGATGCGCTTAAATCATGACCTTTTGGCTACCTGTC 49645  
 Qy 1378 AGCTTTGTTATATGACATTTCTTTTGAAGAAATTCAAATATAGATTGCAAGCCATTGG 1437  
 Db 49644 AGCTTTGTTATATGACATTTCTTTTGAAGAAATTCAAATATAGATTGCAAGCCATTGG 49585  
 Qy 1438 TGCTTTAATGACTGTATTTGGGATTCATTAACAATATCTTTGATCATTTTGTAT 1497  
 Db 49584 TGCTTTAATGACTGTATTTGGGATTCATTAACAATATCTTTGATCATTTTGTAT 49525  
 Qy 1498 TCCGGAAGATCGCAAGGAACTGTTTAAACCATGATGATTTAGTAAATGATGCGCT 1557  
 Db 49524 TCCGGAAGATCGCAAGGAACTGTTTAAACCATGATGATTTAGTAAATGATGCGCT 49465  
 Qy 1558 TCAAAAGACGTTACGCGCAAGGATATGACAAACAGCTACAACTATCAGTTTGTAAAT 1617  
 Db 49464 TCAAAAGACGTTACGCGCAAGGATATGACAAACAGCTACAACTATCAGTTTGTAAAT 49405  
 Qy 1618 GCTTTTGTATATGAGCGGCTCTCTGTCTTTAATTTTGTGATTTATATGACCAATAGGAT 1677  
 Db 49404 GCTTTTGTATATGAGCGGCTCTCTGTCTTTAATTTTGTGATTTATATGACCAATAGGAT 49345  
 Qy 1678 TCTTCTAGAACTTATATGCTCTCTTATATATGACCAACCTGCTGTTTATATGATGCG 1737  
 Db 49344 TCTTCTAGAACTTATATGCTCTCTTATATATGACCAACCTGCTGTTTATATGATGCG 49285  
 Qy 1738 TAAAGAAATCGCTCAAAATATAGTACGTTAACTTAACGTTAGCAATATAA 1797  
 Db 49284 TAAAGAAATCGCTCAAAATATAGTACGTTAACTTAACGTTAGCAATATAA 49225  
 Qy 1798 ATCTCTTTGGAGCTTATAGTCCCAAGGCGCTGTGTATTAATTTATGCAAAATTCAG 1857  
 Db 49224 ATCTCTTTGGAGCTTATAGTCCCAAGGCGCTGTGTATTAATTTATGCAAAATTCAG 49165  
 Qy 1858 ATATATGC 1864  
 Db 49164 ATATATGC 49158

RESULT 3  
 AE002179

LOCUS AE002179 12198 bp DNA linear BCT 30-MAY-2000  
 DEFINITION Chlamydomophila pneumoniae AR39, section 15 of 94 of the complete.  
 ACCESSION AE002179 AE002161  
 VERSION AE002179.2 GI:8163380  
 KEYWORDS  
 SOURCE  
 ORGANISM Chlamydomophila pneumoniae AR39.  
 Chlamydomophila pneumoniae AR39.  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
 REFERENCE  
 AUTHORS  
 1 (bases 1 to 12198)  
 Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F.,  
 White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Uetzerback, T.,  
 Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,  
 Bowman, C., Dodson, R., Gwin, M., Nelson, W., DeBoy, R., Kolonay, J.,  
 McClarty, G., Salzberg, S.L., Eisen, J., and Fraser, C.M.  
 Genome sequences of Chlamydia trachomatis Mogen and Chlamydia  
 pneumoniae AR39  
 Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
 JOURNAL  
 MEDLINE  
 PUBMED 20150255  
 10684935  
 REFERENCE  
 AUTHORS  
 2 (bases 1 to 12198)  
 Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F.,  
 White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Uetzerback, T.,  
 Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,  
 Bowman, C., Dodson, R., Gwin, M., Nelson, W., DeBoy, R., Kolonay, J.,  
 McClarty, G., Salzberg, S.L., Eisen, J., and Fraser, C.M.  
 Direct Submission  
 Submitted (01-MAR-2000) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA  
 On Jun 1, 2000 this sequence version replaced GI:7189108.  
 COMMENT  
 FEATURES  
 source  
 1. 12198  
 /organism="Chlamydomophila pneumoniae AR39"  
 /strain="AR39"  
 /db\_xref="taxon:115711"  
 /note="synonym: Chlamydia pneumoniae AR39"  
 145. .1521  
 /gene="CP0178"  
 145. .1521  
 /gene="CP0178"  
 /note="similar to GB:L26051.SP:P33986.PID:415662;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="UDP-N-acetylglucosamine  
 1-carboxyvinyltransferase"  
 /protein\_id="AAF38052.1"  
 /db\_xref="GI:7189109"  
 /translation="MOIAOVFGCGRLNGEVKVSAGKNAATKLVAISLSDCKTLRV  
 PDIGVSLVETLCKSLGAVSVSDKTEVLEIYETPEIICQTRVPTSPNSVNRIPILILGA  
 LILGCPREGVYVPTVGDAIGERTLTFHREGKQLQVOIISDSGGYAKAPRGKNGYI  
 HLPYPSVATENLILAAIHAKRTVIVKVALEILDLVLFLOKAGADITDNDRTID  
 IFTGGLGSDVHTIIPDKIEASFGMAAVSGRFEVRAKQELIPLKMLRSIGGG  
 FLVSGGLEFFQERLVGVGVLETVHFGFLDMQOQFVAILSDQSSVIEHTVEN  
 RLYVHGLQHMAGBCOLPHQICSTACRAIENPNSAVITHATPLMASHVILPIDRA  
 GRAYVMAHLIAEGSSIIENHILDRGTYNWKRLSLGAKIQIITDMEBEELITPESK  
 LAKRDASL"  
 complement(1475. .3160)  
 /gene="CP0179"  
 complement(1475. .3160)  
 /gene="CP0179"  
 /note="similar to PID:1001350.PID:1001346; identified by  
 sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="arginyl-L-cRNA synthetase"  
 /protein\_id="AAF38053.1"  
 /db\_xref="GI:7189110"  
 /translation="WSTLSTLSVTCQAIKAFPLNEMDAEITPSTKEHGYOCN  
 DAMKLARVLKAPRAIAEAI VALPEQEPSELEIAGAGINTFSVPVFNQDLNHEKD  
 ALKGFVQSPKKRIIIDSFPNIADQMVGHLRSTIIGDSLARIFSYVGHQDLRLNHI  
 GMWGTAIFMLITYLOENPCYDLEDLTSLYKAYVCFNDEPKRSQONVVALQAK  
 DPQALAIWEKICETSEKAFQKIYDILDIYVEKRSFVNPFLPEIIEDEKKGGLTIVS

NDACVPHFAFSPFMVQKSDGGVYATTDLAAMRYRIEDHADKIIIVTDLQSLHF  
QLLEATAAGYLQPGFHSVHGLVLDPOCKLKTSGENVKLRLELLDTAIEKAEAA  
LREHPEIATRAPAVTIGINAIKYSDLSHSRTSDYFVSFKMLRFEQNTAMFLLY  
AVYRIQIGIRLIGISQLSGPPEIQSPABELLATLRRFEALESTIKELCHFLTD  
YLVNLTKFNGFRDSDHQSPYAKSRFLCALAEQVLATCMHLLGLKTLERL"  
complement (3165. .3803)  
/gene="CP0180"  
complement (3165. .3803)  
/gene="CP0180"  
/note="1-acyl-sn-glycerol-3-phosphate acyltransferase,  
putative; identified by match to PFAM protein family HMM  
PF01553"  
/codon\_start=1  
/transl\_table=11  
/product="1-acyl-sn-glycerol-3-phosphate acyltransferase,  
putative"  
/protein\_id="AAF73638.1"  
/db\_xref="GI:8163381"  
/translation="MIFRICKFTFWVAFSLFYKLVGVKQNFIKGPAIIAVNHSPL  
DPIALHMCVHECIYHLARSLFNIPWLKQMGCFVPRQDEGNSAAFKIASRLFNKRX  
LVYPEGASPDQGPQKVGIGMMAAKSRVPIIPVYIRGTFFAFNRHQKIPHWKTI  
TCVFTPMYFDIIQNEIKNKETYQIITQTNKIAELKAWYESGCKGVP"  
complement (3800. .4450)  
/gene="CP0181"  
complement (3800. .4450)  
/gene="CP0181"  
/note="similar to SP:P23863 PID:42839 GB:U00096  
PID:1651431 PID:1651438; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="cytidylate kinase"  
/protein\_id="AAF38054.1"  
/db\_xref="GI:7189111"  
/translation="MLITIDGPGTGTAKALADHLHFNVCNTGKRVETLAVARLQ  
SPWATLPLTLEBPSPFATGQOPLESFNGHLLSELITQEVANAASLSQPEV  
RAFMDIQRRAQGNVCEFRDGMGSKVPFNADLKIFLTSSPEVRAQRRLKDLPEGTL  
SPEQLAELVKRDAADRAQRAHPLVPIENGIVDSSDLTIRQVLEKILALFRNEL"  
complement (4447. .5373)  
/gene="CP0182"  
complement (4447. .5373)  
/gene="CP0182"  
/note="similar to GB:ML1330 SP:P06466 PID:1208947  
PID:145476 GB:U00096; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="phosphatidate cytidyltransferase, putative"  
/protein\_id="AAF38055.1"  
/db\_xref="GI:7189112"  
/translation="MLNSNPKSKTGAYGDLFORVVVHSLVLTPLVLLLYSSLPPLTS  
FALGFIATCGAVCTEYSSMAKAKHYPLSTFSAIGSFLFLALSFLISWGHSLPGF  
FDALPWTLLIIVWVMSIFPRVKSTIGALQISGTLFSLIVGIPRILFLVHVSFIHT  
QEPVLGWMASFLIATGADIFGFFGKAFGNKKIAPQISPNKTVGVFVAGCGIAT  
ISFIFIQIIPRFASYFPMFPAIPLGLALGITGFGDIIIESIFKRDHALKNSNKLKA  
VGMGLDLSLLSTPIAYLPLITQSKERIG"  
complement (5374. .6159)  
/gene="CP0183"  
complement (5374. .6159)  
/gene="CP0183"  
/note="similar to PID:1208946 GB:U00096 SP:Q47675  
PID:1552751 PID:1786371; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="undecaprenyl pyrophosphate synthetase"  
/protein\_id="AAF38056.1"  
/db\_xref="GI:7189113"  
/translation="MTVALINDPVDLSLATNNAESKPSLQRLPNHVAIIMDGNRWY  
KKHRECGHHTSGHYGAKVLPNILNAVLDLGLKVLTLTYTFSTENGPRKPEIQEIF  
NIFYTQDKQPLWMEINEICLRGIDLSKLPKGIQTKINHVRNMTASFSRLELVLAV  
YGGDELIVRAFKLHVDILNKISLSSLDSTSLYDITGLDTPDLLIRITGEMRVS  
NFLWQIAYTELYITDLPDFTPDQLFEAINVYQSRRRGK"

complement (6214. .6363)  
/gene="CP0184"  
complement (6214. .6363)  
/gene="CP0184"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF38057.1"  
/db\_xref="GI:7189114"  
/translation="MPKKPLYSFEVRTLKNFLLETPICLNRIQPLPTGMAGEKSFVI  
FVLGK"  
6637. .7737  
/gene="CP0185"  
6637. .7737  
/gene="CP0185"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF38058.1"  
/db\_xref="GI:7189115"  
/translation="MKPFRISDLIPTQMVIWRGGNVHYVPAQNLPKKILGGVL  
ACFGLALLGCAFAAGVQCOTIFPCIGLMILGLVLLGAYLQYSKGRFRPLPRETK  
VFEBPINWLGLSLLQSWKKIRPGCVYHPCPOVEICEGSOEIVTKI FOKKSDRNTSI  
FLIQEMDOIALRGIEKSSLSRKTFAIDPSVSSLLSEIOREEQOVLDPKVISWSSD  
QASDRTHPKSAIYVNISDAQAEPQRCYIDAYTKAFTVLDQIDQIDPNIVKHTIYVLT  
PDLRTHALPKBEQENLKLLSQAAFYSQAQVAKRMREKQDSIRIKFIPTDPTSPTS  
LYFSPHSSPTPHSVTPISLGFVGEQESYTA"  
7917. .12125  
/gene="CP0186"  
7917. .12125  
/gene="CP0186"  
/note="similar to GP:3220156; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="secDF protein, putative"  
/protein\_id="AAF38059.1"  
/db\_xref="GI:7189116"  
/translation="MKQVKRNFALLICVFALALYYVLTCLYYAKPLDKKIDGNEAE  
HIKSTQKAAQVRKDVIPRVSALLSLHLRHGHIQHPAIPDIVSVVRFKGEDAEDFI  
GNLVHGEPNVPIKSARLVGVYGREHDDHVIQVASSINTSLVSDSFVSYSSENEOE  
MASSILQRYVSACTFPKQDCSCSYIMETAPEKQLLOYAKNLSGSEFVPSRSLAP  
CQGSFSDQLRAFLRSLSDSLNDAALVDEQDLKLSVETLSQTACIRLDCDFYIEG  
LRJDCSESLFFSSIEYCPKERKIFLTHSDLLAQRSTLSKEQRLDPDSLRLAVEKQL  
IPENFPVFCRQPRESEAFGCYIFSPNTCDKHFSKGVYILKGLRSTIVAKYQGGGKE  
LOSFEKDLQNLVNCFSHTEAISWTLGEDOVLEIRHPLQOFLDVMGEGFVIGKEGCAF  
EVXDIQDRLATVNIQIKNRQSDLVVRHEQVYRHAKCSMDLQERLSAPITYONLFLNNK  
LNNKFRSGENILRLGLDPVGGKQLLSFKNDHQKQLTDKEDILKVSDELCARLINKLG  
VSBIEURREGDYIHLSPVGSSTISSSEILGTSTKMSPHVNVNERFSSYSASRYEQRFLD

Query Match 96.4%; Score 1797.4; DB 1; Length 12198;  
Best Local Similarity 99.8%; Pred No. 0;  
Matches 1831; Conservative 0; Mismatches 1; Indels 3; Gaps 3;  
Qy 1 ATGCACTTCGCATATTGTTCAGGAGGGGATCAGCGGCACCTGCTAATGACAATATTTCTGC 60  
Db 10364 ATGCACTTCGCATATTGTTCAGGAGGGGATCAGCGGCACCTGCTAATGACAATATTTCTGC 10423  
Qy 61 AAACCGTGGATGGCGTATGGCTGTAGTGATGACGGTTATATGTCAGCACCCCTATTTT 120  
Db 10424 AAACCGTGGATGGCGTATGGCTGTAGTGATGACGGTTATATGTCAGCACCCCTATTTT 10483  
Qy 121 AAACGTCCCATTTGAAAATCATCGACGTCTCAGGGAATTTTACCCACCGTGAAGTCAG 180  
Db 10484 AAACGTCCCATTTGAAAATCATCGACGTCTCAGGGAATTTTACCCACCGTGAAGTCAG 10543  
Qy 181 CAAACTCGCCTCAGATTAAATCTGGAGCGATGCTCTTTTGTCCCGAGGTTCTCAGTGA 240

Db 10544 CAAACGCGCTCAGATTTAAATCTGAGCGATGCTTTGTGCCGAGGTTCTCAGGA 10603  
 QY 241 AGAGACGATCTCTTCTGATCTTGGAAAAAACAATGACACAGGCAATTATCTCAGATG 300  
 Db 10604 AGAGACGATCTCTTCTGATCTTGGAAAAAACAATGACACAGGCAATTATCTCAGATG 10663  
 QY 301 CTGAGGCTGGCAATGCTTATGTTTGAAGAGCTATATATGATTTGAGAGCGTCAT 360  
 Db 10664 CTGAGGCTGGCAATGCTTATGTTTGAAGAGCTATATATGATTTGAGAGCGTCAT 10723  
 QY 361 CGCTTCGGAGCTGTTCTTCTGAATCTTTGCTTATCTGGGAGCTCAGATATTTGGA 420  
 Db 10724 CGCTTCGGAGCTGTTCTTCTGAATCTTTGCTTATCTGGGAGCTCAGATATTTGGA 10783  
 QY 421 TGGCCACACACCTTGTGAGGACTGCTGGAGTTGTTCTGTATGAGGATGCGGTGA 480  
 Db 10784 TGGCCACACACCTTGTGAGGACTGCTGGAGTTGTTCTGTATGAGGATGCGGTGA 10843  
 QY 481 TGCAGATGTTCTGTATCTGAGGAAATCCGAGAGAAATTTTATGTTCTCAAGCTTAA 540  
 Db 10844 TGCAGATGTTCTGTATCTGAGGAAATCCGAGAGAAATTTTATGTTCTCAAGCTTAA 10903  
 QY 541 AAAATCTGTAGAAAAAGATATACCAAGGCTTTGAGGCAATTTTGAATCTCACTTAC 600  
 Db 10904 AAAATCTGTAGAAAAAGATATACCAAGGCTTTGAGGCAATTTTGAATCTCACTTAC 10963  
 QY 601 TACAGTATGAGCTCAGGACTCTTTCTTCTAGATACAGGCGCTATTTAAAGGCTTTC 660  
 Db 10964 TACAGTATGAGCTCAGGACTCTTTCTTCTAGATACAGGCGCTATTTAAAGGCTTTC 11023  
 QY 661 TTTGACATGATTTTGAAGATTTTCTCTCAATGTTTACGGCTCTTTTCAAGCAATTA 720  
 Db 11024 TTTGACATGATTTTGAAGATTTTCTCTCAATGTTTACGGCTCTTTTCAAGCAATTA 11083  
 QY 721 TTTCTCAATGCTGTGATGATTAAGACCAACATACAGTGTGATGATTAAGT 780  
 Db 11084 TTTCTCAATGCTGTGATGATTAAGACCAACATACAGTGTGATGATTAAGT 11143  
 QY 781 CGTGGGATTAAGATGATTTCTTGAGAGATGCAAAAAAATCTTGGGCTGTTCTGAGG 840  
 Db 11144 CGTGGGATTAAGATGATTTCTTGAGAGATGCAAAAAAATCTTGGGCTGTTCTGAGG 11203  
 QY 841 TGTCTTCTTTTGAAGTGTGCTCTCTGCTGGGTTTGAAGCTGGAATTCGCTTGGGAT 900  
 Db 11204 TGTCTTCTTTTGAAGTGTGCTCTCTGCTGGGTTTGAAGCTGGAATTCGCTTGGGAT 11263  
 QY 901 GGATTTTAAAGAGGATGCTTACCTTAAATCCAAAAGAGATGAGCATCAGGATGT 960  
 Db 11264 GGATTTTAAAGAGGATGCTTACCTTAAATCCAAAAGAGATGAGCATCAGGATGT 11323  
 QY 961 TGTCTCAATGCTGTGCAAAAGTTGTGATTAACATACAGGAGCTGCTCTTCTTGAAGA 1020  
 Db 11324 TGTCTCAATGCTGTGCAAAAGTTGTGATTAACATACAGGAGCTGCTCTTCTTGAAGA 11383  
 QY 1021 CTTCCGATTTCAAAACATTTGATCTTCAAGAAAAATCAATCTATTTTATGATTAAGC 1080  
 Db 11384 CTTCCGATTTCAAAACATTTGATCTTCAAGAAAAATCAATCTATTTTATGATTAAGC 11443  
 QY 1081 TTTAAGCTATCT- AAGCAGATACAGGCTCTCTCTTAAATTTAAGATCAATGAGCTGCG 1139  
 Db 11444 TTTAAGCTATCTAAGCAGATACAGGCTCTCTCTTAAATTTAAGATCAATGAGCTGCG 11503  
 QY 1140 GTTA- TTGTGGATTTGTGCAAAACAGGCTTAAATTTCTTACGAACTCTTAAAGCA 1198  
 Db 11504 GTTA- TTGTGGATTTGTGCAAAACAGGCTTAAATTTCTTACGAACTCTTAAAGCA 11563  
 QY 1199 AAGCGAAA- TTTGTGTAAAGGATTAAGCAAAATATCAAGAAAGAAAGCGCTTACAGG 1257  
 Db 11564 AAGCGAAA- TTTGTGTAAAGGATTAAGCAAAATATCAAGAAAGAAAGCGCTTACAGG 11623  
 QY 1258 GACCATGGGCTTTTGAAGAGCTTGGCAATCATCTGCTATGATGATTTGGCTTGA 1317  
 Db 11624 GACCATGGGCTTTTGAAGAGCTTGGCAATCATCTGCTATGATGATTTGGCTTGA 11683

QY 1318 ATGGCAATATGCTTTTCAAGTCCGTATGCGCTTAATTCATGACCTTTTGGCTACCTGTC 1377.  
 Db 11684 ATGGCAATATGCTTTTCAAGTCCGTATGCGCTTAATTCATGACCTTTTGGCTACCTGTC 11743  
 QY 1378 AGTCTGTATATGACATTTCTTTTGAAGAAATTTCAATATGATTTGCAAGCCATTTGG 1437  
 Db 11744 AGTCTGTATATGACATTTCTTTTGAAGAAATTTCAATATGATTTGCAAGCCATTTGG 11803  
 QY 1438 TGCTTATATGACGTATATGAGGATTTATCATTTAAACAATCTTTGATCATTTTGTATG 1497  
 Db 11804 TGCTTATATGACGTATATGAGGATTTATCATTTAAACAATCTTTGATCATTTTGTATG 11863  
 QY 1498 TCGTGAAGATCGCCAGACCAACCTGTTTACCCCTATGACATGTTTATGATGACCT 1557  
 Db 11864 TCGTGAAGATCGCCAGACCAACCTGTTTACCCCTATGACATGTTTATGATGACCT 11923  
 QY 1558 TCAAAAGACGTTACGCGGACGCGGTAATGACACAGCTACACTCTATCAGTTTGTAA 1617  
 Db 11924 TCAAAAGACGTTACGCGGACGCGGTAATGACACAGCTACACTCTATCAGTTTGTAA 11983  
 QY 1618 GCTTTGTATATAGCGGCTCTCTGTCTTAAATTTTGAATTTATGACATAGGAT 1677  
 Db 11984 GCTTTGTATATAGCGGCTCTCTGTCTTAAATTTTGAATTTATGACATAGGAT 12043  
 QY 1678 TCTTCTAGAACCTTATCTGCTCTTATATTTGACACCACTCTGTTGTTATGCTCCG 1737  
 Db 12044 TCTTCTAGAACCTTATCTGCTCTTATATTTGACACCACTCTGTTGTTATGCTCCG 12103  
 QY 1738 TAAAGAAATCGCTCAAAATATAGTACGTTAACTTATCTAAGTATGACATATPAAA 1797  
 Db 12104 TAAAGAAATCGCTCAAAATATAGTACGTTAACTTATCTAAGTATGACATATPAAA 12163  
 QY 1798 ATCTCTTTGGGACTTTAGTCCCAAAGCCCTGT 1832  
 Db 12164 ATCTCTTTGGGACTTTAGTCCCAAAGCCCTGT 12198

RESULT 4  
 AE001318/c 9989 bp DNA linear BCT 30-OCT-2000  
 LOCUS  
 DEFINITION Chlamydia trachomatis section 45 of 87 of the complete genome.  
 ACCESSION AE001318 AE001273  
 VERSION  
 KEYWORDS AE001318.1 GI:3328875  
 SOURCE  
 ORGANISM Chlamydia trachomatis.  
 Chlamydia trachomatis  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 REFERENCE  
 AUTHORS Stephens, R.S., Kaiman, S., Lammel, C.J., Fan, J., Marathe, R.,  
 1 (bases 1 to 9989)  
 Aravind, L., Mitchell, W.P., Olinger, L., Tatunov, R.L., Zhao, Q.,  
 Koonin, E.V. and Davis, R.W.  
 Genome sequence of an obligate intracellular pathogen of humans:  
 Chlamydia trachomatis  
 JOURNAL Science 282 (5389), 754-759 (1998)  
 MEDLINE 9900809  
 PUBMED 9784136  
 REFERENCE  
 AUTHORS 2 (bases 1 to 9989)  
 Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,  
 Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.  
 Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
 JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
 MEDLINE 99206606  
 PUBMED 10192388  
 REFERENCE  
 AUTHORS 3 (bases 1 to 9989)  
 Stephens, R.S., Kaiman, S., Lammel, C.J., Fan, J., Marathe, R.,  
 Aravind, L., Mitchell, W.P., Olinger, L., Tatunov, R.L., Zhao, Q.,  
 Koonin, E.V. and Davis, R.W.  
 Direct Substitution  
 JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University  
 of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA  
 FEATURES  
 source 1..9989



```

/organism="Chlamydia trachomatis"
/strain="UW-3/Cx"
/db_xref="taxon:813"
complement(105..371)
/genes="omcA"
/note="CT444"
complement(105..371)
/genes="omcA"
/codon_start=1
/transl_table=11
/product="9kDa-Cysteine-Rich Lipoprotein"
/protein_id="AAC68043.1"
/db_xref="GI:3328876"
/transl_table="MKKTALLAALCSVVSLSSCCRIVDCFBPDCAPIQCSPECSSKK
DVGCGNSCGYVPACKCGDTHQDANGFQARGIPVDGKCRQ"
708..932
/genes="CT444.1"
708..932
/genes="CT444.1"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC68048.1"
/db_xref="GI:3522897"
/transl_table="MKRFELFIVATAALLTQGSYBPCKREYILAQKCSPLFKHAQTL
DFSQAQVEMTFTFKAPPCVSSLELSQN"
complement(929..2449)
/genes="gltx"
/note="CT445"
complement(929..2449)
/genes="gltx"
/codon_start=1
/transl_table=11
/product="Glutamyl-tRNA Synthetase"
/protein_id="AAC68044.1"
/db_xref="GI:3328877"
/transl_table="MTQNVVRVAPSGTGDPHVGTAYMALFNEVPARKYNGOMILRI
EDIDQTRSDYANFTSALUKWCIRWDEGPDVGGAGPYRQSERTEIYKKAIEILLQ
TDCAYKCFATPQBLQENRAVASTGLYRGYDRYRLSPVEEVEEQGPYTRILKY
PLTGESVFDQCGCVVFWADVDQVLVKSDFPTTFHFANVVDHLMGITHVLGRGE
WLSSTPKHLLLYEAFGPEPPQFPHMLPDGSKLSKRNKPTSIFFYRDAGYKGEAF
MNFLLTLMGYSMEGDEITYSMORLIEADPKRIGRSGAVPDIRKLDNMNKHVLANHEGSP
ESILQELKGLHWDEFLLKILPICOSRITTLADPVLGTSFFFTAIPOYSKRELLPSGL
KQSQAAVLYSLVKYLEKDLWEDKPFYQGSKWLAEAFVHHKKAIVIPLYVAITGAK
QGLPFDMSLELGGKARTARLTVAQNLLGVSVKVQQQVDKALQDQPLEDIRLDF"
complement(2721..3272)
/genes="euc"
/note="CT446"
complement(2721..3272)
/genes="euc"
/codon_start=1
/transl_table=11
/product="CHLPS Euc Protein"
/protein_id="AAC68045.1"
/db_xref="GI:3328878"
/transl_table="MECLQDQTVGEABQVQVQOQENAVPVTSQRVSIQAAKLHNV
RQAYVAIKOKLKASKTRWEIDLOLEDYRNRYSRAKSTYQGLLFDNERGFYSV
QGVASMDVPEQIKYIYATRIGAMKGERGSAWIHVSEVDRYRNDYLKKEARKGKSL
AAMREGFALGADLLADAEAFIS"
complement(3677..5431)
/genes="recJ"
/note="CT447"
complement(3677..5431)
/genes="recJ"
/codon_start=1
/transl_table=11
/product="ssDNA Exonuclease"
/protein_id="AAC68046.1"
/db_xref="GI:3328879"
/transl_table="MEKDSLSLQKWDYPKHNEVFKTLKEFHLHPATAQVLIISRG
FQSIQETRDFLPQLSLHSTSLFLDMEKAVARLLOAKANNEHVMIYDGDVDGIGTV
TLIVEFQILQVGTSCCSGTLFKOGETASLISQMLQDGISLITIVDCCGITAGKEVH
AINKQGIIDVITVDHMTGKPLPHCVAMLPKLDKNPYPNKELTGVGVAFKLVCATYEE

```

```

LIQODASWKDKIDLLRFLDLSVTIADVGRSGENRILVSYGIEKIAKGRGLGLKL
CSLSGVDSKSEVSTNLGIRITPKLNSGLRLADSSQGVKLLSLHSDHPNNIGTIVSELSAV
NOERORIEAEVLDRVERILAAANPKLTAQSAIVLASPNWHSRVPIIISARLARTYKPV
AIALQDGIKGLSGTIGSPFLGLVRKESFPLSYGCHDFAAGLMIKEDQVGGFRKK
FIHLVSSSLRKDDAMRTLSLDVMDPSRINRDLIASMELLEPEFGKGNVSFVYTKAIO
VRYPKLAGNHVKLYLNSGERNLEGTAGGQDKISULKANMNIPLDIATVTLRMRSA
RGAIRLLIQDFRIQIPRL"
complement(5549..9751)
/genes="secD/secF"
/note="CT448"
complement(5549..9751)
/genes="secD/secF"
/note="fusion"
/codon_start=1
/transl_table=11
/product="Protein Export"
/protein_id="AAC68047.1"
/db_xref="GI:3328880"
/transl_table="MKENFRNLSIILVFPALAHVLPCTCLYYSRPLNRKIDGREAR
KIERETROVSARDNLTURVAKILSTLKLKRGDIKOHPSIPGVNVHFFPLAEADAALFI
DNVHGEPVPVKSRARLYVSGSBFTGETVQVIGSLVTSLEKDSFVSYEDKQTVFV
SOBLELATNLFIESERPCEGYSGLWDSMPIKVIYVTRNICSLRLPLPARKTQPLV
SRCLGTERDFTAPLARLEKALOHEEVKEAREELKQASYLHLSNRNLRWNSISPKVGVN
VDCSELSPCESMSLTADCKLPHEDAENANRQOLVGNERLDLERLFAKEOHIATR
LARSVEGTEFGFAVRLRDEAASGKIVLGRVRCQRIVEHLTALVLRNPINAECDLSTE
NFPVYREPLDSDALGCFVSPERSCRHFSKGSVIYFKGLRSIVAKYEHAGAEAL
LQHDQLNYACFIHTDASVSLGEDRVLEIKEPLORVIRIWGEFVQSGKASLEVRD
VRDLAVNRIEKTQHAELVRWDEQROQCSNPRVLRRAAI PHKNIIPENLKNIR
KYSRSHVLRFGTDFVGGKQIRIARFDHGNLLTDKAGIDKVSDELVARLUNKLVSEV
GMOREGDHLOVSPVGAAGISSADILGTSKMSHVNEQPSRSPLEYQVTFIDYLWF
TARSLDECSPOAINRLAGFHDNDGASAPNVRVAVEKREAGLSFSKELEGGASLD
TOYSMTAIEKESREQVNLPMI VFRNHALEGASLNRIPBFAVGEVYVGNFVXDKATF
SDRETVPQVQFHAMTSKFCQBGSGTNGFLFSGRCGRMAVLDGVVIDPVLNVPLK
DHASVSGNFSYREVRHLAALDKSGAMSFPEILSEBVSPELSSORVOQILSVLGL
VLIIVLMSYVFRGVIASIVALLNLLIWSQYLDAPLTLSGLAGIILAMGWADA
NVLVFERIEEYLLTFESLSEVEGKAFSAIEDNLATILASALLMLDTCPIKGF
ALTLIGISFSSMFTALPMTKFPFVHVKTRTQTLHMMKFGIKKINFLKECRLLWV
SGSVLVGCVGLGFGAWDFGMDFKGVALTLDSDVCYNPEQMSVLURKRFQOIGL
SRDYRFRADSESEKVIYLSNALDRVQIEGASEQKSDYHSLARVQLVSDSGSS
TTSWMDPADSGWFSVGGOLSNKMTQAVIALFGALGILFVYSLRPEWRYAFSAICA
LMHDLATCAVLVALHFFLQRIQIDLOAIGALMTVLGYSLNNTLIIIPDRIRDRREKL
FTPMPLINDALQKTLGRVMTATTLSVLVILLFVGGGSI NFNAFTMTVGTILLGTLIS
SFIAPPLULFVVRKEEQNSLR"
BASE COUNT 3118 a 2399 c 1722 g 2750 t
ORIGIN
Query Match 41.5%; Score 773.6; DB 1; Length 9889;
Best Local Similarity 66.8%; Pred. No. 1.5e-161;
Matches 1170; Conservative 0; Mismatches 574; Indels 12; Gaps 4;
Qy 2 TGGACTTCCGCATATTGTCAGAGGGGATCAGCGGCACTGCTAATGGACAATATTCTGCA 61
Db 7318 TGGACATCAAGITTTGCCAAGAAGGGTTAGTGGTACAAAAATGCGCTTTCTCTGGA 7259
Qy 62 AACCGTGATGGCGTATGGCTGTAGTGAITTAGCGGTTATATGGTCAGAGCCCTATTTTA 121
Db 7258 GGGAGAGGGTGGCGTATGGCTGTATAGATGGCTACCGTATCAGCGATCCCGTGTG 7199
Qy 122 AACGTCCCATTTGAAAATCATGCCAGTGCTCAGGGGAAATTTACCCACCGTGAAGTGAGC 181
Db 7198 AATGTTCCACTGAAAGATCATGCTAGTGTCTTCTGGAACATTTCTCATACAGAAAGTACAT 7139
Qy 182 AAACCTCGCCTCAGATTTAAATCTGGAGCATGTCTTTTGTTCGAGGTTCTCAGTGAA 241
Db 7138 CGTTAGCTGAGATCTTAATCTGGAGCTATGTCTTTCATCCAGAAATTTTAAGTGAA 7079
Qy 242 GAGACGATCTCTTCTGATCTTGGGAAAAAACAATGTACACAAGGCATTATCTCAGCATGC 301
Db 7078 GAAGTGGTATCTCTGAGTTGGGAAGTTCTCAGCGTGTACAAGAAATCTCTCTGTTGTG 7019
Qy 302 TGTGGCTTGGCAATGCTTATTTTGTATGAGCGTATATATAGATTTGGAGGCGTCATC 361
Db 7018 TTGGGACTTGTAGTGTCTCATCGTAACTAAGTGTGTACTACAGATTTGGAGGCTGTATT 6959

```

QY	1433	ATTGTCGTTTAATGACGTCATTTGGGGATATTCATTAACAATACCTTGATCATTTTGGAT	1492
Db	5881	ATGAGCGCATTTAGACAGAGTCCTAGGGTACTCTTGAACAATACGTTGATTAATTTTGGAT	5822
QY	1493	CGTATTCGTGTAAGATCGCCAAACGACACCTGTTTACCCCTATGATGTTTATAGTAATGAT	1552
Db	5821	CGTATTCGTAAGATTCGGCGGAAACAACTGTTTACCCCAATGCGATTTGGATCAATGAT	5762
QY	1553	GCCCTTCAAAGACGTTACGCCGACGGTAAATGACAACAGTAACTATACGTTTGG	1612
Db	5761	GCATTGACGAAGACATTAGACGTPACAGTAAATGACTACGGACAACTATCTGTACTG	5702
QY	1613	TTAATGCTTTTGGTTTATAGGGCGGCTCCCTGCTCTTAAATTTTGCATTATATAGCAATA	1672
Db	5701	GTCATTTTGTATTATTCGTTGGGGAGGTTCTATTTCATTTTGCATTATATAGCAGTG	5642
QY	1673	GGGATTCCTTAGAACAATTATATCGTCTCTTATATATGACCAACCTCTGTTGTTTATG	1732
Db	5641	GGGATTCCTTAGAACAACGCTATCTTCAATTTGACATAGCTCTCCGCTTCCTTATATGAC	5582
QY	1733	GTCGCTAAAGAAATTC	1748
Db	5581	GTGCGTAAAGAAAC	5566
RESULT 5	AE002341/c	12442 bp	DNA linear BCT 26-MAY-2000
LOCUS	AE002341	12442 bp	DNA linear BCT 26-MAY-2000
DEFINITION	Chlamydia muridarum, section 69 of 85 of the complete genome.		
ACCESSION	AE002341	AE002160	
VERSION	AE002341.2	GI:8163305	
KEYWORDS			
SOURCE	Chlamydia muridarum.		
ORGANISM	Chlamydia muridarum		
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.		
AUTHORS	1 (bases 1 to 12442) Read, T.D., Brunham, R., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utechtback, T., Berry, K., Bess, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., Deboy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.		
TITLE	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AK39		
JOURNAL	Nucleic Acids Res. 28 (6), 1397-1406 (2000)		
MEDLINE	20150255		
PUBMED	10684935		
REFERENCE	2 (bases 1 to 12442) Read, T.D., Brunham, R., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utechtback, T., Berry, K., Bess, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., Deboy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.		
AUTHORS	Direct Submission		
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712		
JOURNAL	Medical Center Dr, Rockville, MD 20850, USA		
COMMENT	On Jun 1, 2000 this sequence version replaced gi:7190754.		
FEATURES	Location/Qualifiers		
source	1..12442 /organism="Chlamydia muridarum" /db_xref="taxon:83560" /note="synonym: Chlamydia trachomatis MoPn" complement (86..544) /gene="TC0726" complement (86..544) /gene="TC0726" /note="similar to SP:P18587; identified by sequence similarity; putative" codon_start=1 /transl_table=1 /product="15 kDa cysteine-rich outer membrane protein" /protein_id="AAP39536.1" /db_xref="GI:7190755"		









<p>RLIHONVLEFDEKLGKRPVLIICHGVSWNEEYORGILNVIKNHLSGNSAILSTYNTY  PGKRNLEVARDDWLFREDENLKNRGEQINSENAVYKRGALRPIESQSIINERKACIN  GITEKDDVYLHLEFENNNPLYIDNKKLLEGLIHVDSJLMTKFPNLSNREK  LSAECQNDJIAKEQYDFLDRQFRISIVHANKKINISKVRITLDEKIDIRKY  OKKKGDFYIENNEIKDEEISLIDILISENPVLTIDELEKREKKNLENNVAN  AVYKDGKLEAVSRKLVKEEKIKLNSKYQYLVNPIIDENPVIALASYEETVAD  NENPMFLMTFLDGRTRTDEDINLLEKKEBVEVITPESSKEEVIKNNEICGN  PIEINPLNF"</p>		gene	complement (10436..10513) /gene="FN0702" complement (10436..10513) /gene="FN0702" /EC_number="3.5.1.14" /codon_start=1 /transl_table=1 /product="N-acyl-L-amino acid amidohydrolase" /protein_id="AAL94898.1" /db_xref="GI:19714229" /translation="MDEALFMGANLYAQFAIDPLNSK"
<p>complement (10579..11235) /gene="FN0703" complement (10579..11235) /gene="FN0703" /EC_number="3.5.1.14" /codon_start=1 /transl_table=1 /product="N-acyl-L-amino acid amidohydrolase" /protein_id="AAL94899.1" /db_xref="GI:19714230" /translation="MENCYINSWYFGNKAKATIEESKIIDSIDTTFALHMQGVV  GKTSISGAMMAADLPSIKVKGSGHSPHETIDVAVASAIVMNLOHLVARNTP  LDTLVTVVKGTLTRNNTIAGEMLLEGTTRTSRDEPWKRVPSQIEVYKNTAAVDA  EVENLVRATPLVNDQDISNLTSAKLYGEBVVTKEKTSGEVHLHLKFLVH"</p>		CDS	complement (11172..11366)
<p>Query Match 8.5%; Score 161.2; DB 1; Length 11527;  Best Local Similarity 54.3%; Pred. No. 1.5e-25;  Matches 325; Conservative 0; Mismatches 273; Indels 0; Gaps 0;</p>		gene	complement (11172..11366)
Qy	148 TGTCTCAGGGAATTTTACACCGGAGTACGACAACTCCGCTCAGATTAAATCTGG 207		
Db	7196 TATACAGGAATTTACATGTTGAAGAGCAAGAGGACACAGATTTAAATGACGG 7255		
Qy	208 AGCAGTCTTTTGTCCAGGTTCTCAGTGAAGACGATCTTCTGATCTTGGAA 267		
Db	7256 GGCAATTGCCAATTAAGCAAGATAGCTGAACAAGACTGTGGAGCTACCTTGAGA 7315		
Qy	268 AAAACATGTACACAGCATTTATCTGCATGCTGTGGCAATGCTTATTTGTTT 327		
Db	7316 TGAATCATACCTCAAAATGAAAATGAGTATGCTGCAATGTTTAAATTTGGGCTT 7375		
Qy	328 GATGAGCGTATATTATAGATTGGAGCGTCATCGCTTCGGAGCTGTTCTTGAATCT 387		
Db	7376 TATGATATCTTTTATATACGAGAAATTAACAGATTTGGCAATATATATTTTGG 7435		
Qy	388 TTTCCTTATCTGGCAGCTCTACAGTATTTGATGCGCCACTCACCTTGTACAGACTGC 447		
Db	7436 ATTATATACATTTGCTGTCTTAAATTTATAGATGCAACACTTACTTACAGGATATGC 7495		
Qy	448 TGGATTTCTTGTCTATGGGATGCGCTAGATGCAATGTTCTTGTATTCGAAAGAT 507		
Db	7496 TGGATTTATCTTTCACCTTGGAATGAGCTGATGCAATGTTTATTTTGGAGAGAT 7555		
Qy	508 CCGAGAGAAATTTTATGCTCAAAAGCTTAAAAAATCTGTAAGAAAAGATATACCA 567		
Db	7556 TAAAGAGAGTTACATTTGAAATATGCAATAAGAAATTCATAGATTCGCTTTAARA 7615		
Qy	568 GGCCTTTGAGCCATTTTGAATCTTAACTGATCAACAGTATTTGGCTCAGACCTCTTT 627		
Db	7616 AGCTTTTATGCAATATTTGACTCAACCTTAACTTAAATATTAACAATCTTATTT 7675		
Qy	628 CTTCTAGATACAGGCGCTATTAAGGCTTTCCTTGACATTTGATTTAGAAATTTTCTC 687		
<p>RESULT 7  AE001567  LOCUS  DEFINITION  ACCESSION  VERSION  KEYWORDS  SOURCE  ORGANISM</p>		gene	complement (894..1961)
Db	7676 TGTATTTGACTACGACCAATAAAGATTTGAGTAACATTGACATTGAGAACATTAGC 7735:		
Qy	688 TTTCAATGTTTACGCGCTCTTTTCACTATGACTAAATTTTCTTCATGCTGTGATGATAG 745		
Db	7736 CTCATGTTTACAGCAATACAGTAACAAGATTAATCTTTGACTTTGTGAATATAG 7793		
<p>REFERENCE  AUTHORS  MEDLINE  PUBMED  REFERENCE  AUTHORS  TITLE  JOURNAL</p>		gene	complement (894..1961)
<p>1 (bases 1 to 19836)  Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C.,  Smith, D.R., Noonan, B., Guild, B.C., deJonghe, B.L., Carmel, G.,  Tummino, P.J., Caruso, A., Uria-Nickelsen, M., Mills, D.M., Ives, C.,  Gibson, R., Merberg, D., Mills, S.D., Jiang, O., Taylor, D.E.,  Vovis, G.F. and Trust, T.J.  Genomic sequence comparison of two unrelated isolates of the human  gastric pathogen <i>Helicobacter pylori</i>  Nature 397 (6715), 176-180 (1999)</p>		gene	complement (894..1961)
<p>2 (bases 1 to 19836)  King, B.L., Alm, R.A. and Trust, T.J.  Direct Submission  Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney  Street, Cambridge, MA 02139, USA  Address all correspondence to: hp@arch.us.astro.com or Richard  A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge,  MA, 02139. Jo-See L. Ling, Donald T. Moir, Douglas R. Smith,  Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills,  Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics  Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and  Diane E. Taylor are with the University of Alberta Department of  Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G  2H7 and the Canadian Bacterial Diseases Network. All other authors  are with Astra Research Center Boston, 128 Sidney Street,  Cambridge, MA, 02139. Putative identifications, sequence  alignments, and name and sequence search capability are available  at ARCB's World Wide Web site. (URL:  http://www.astro-boston.com/hpylori).  Location/Qualifiers</p>		gene	complement (894..1961)
<p>FEATURES  source</p>		gene	complement (894..1961)
<p>1..19836  /organism="Helicobacter pylori J99"  /strain="J99"  /db_xref="taxon:85963"  74..715  /gene="hjp143"  74..715  /gene="hjp1443"  /translation="MOKKIFLDEDDVILSESIEKFEHLGVEFCANFGKAEHRLSV  ERNILLVQVDEMSLELFRKIDPLISFVPIITLQDNATKKNFNGASDYL  KRPFDLEFARKREPNDPRLIEMINIVYHQAIVNKKETLAKTAKOLLEFLEH  KGIISQALENNLMEOATDSTLRITIVLRKLKNCIEFHKGVRNPL"</p>		gene	complement (894..1961)
<p>/codon_start=1  /transl_table=1  /product="putative TRANSCRIPTIONAL REGULATOR"  /protein_id="AD07036.1"  /db_xref="GI:4156082"  /translation="MOKKIFLDEDDVILSESIEKFEHLGVEFCANFGKAEHRLSV  ERNILLVQVDEMSLELFRKIDPLISFVPIITLQDNATKKNFNGASDYL  KRPFDLEFARKREPNDPRLIEMINIVYHQAIVNKKETLAKTAKOLLEFLEH  KGIISQALENNLMEOATDSTLRITIVLRKLKNCIEFHKGVRNPL"</p>		gene	complement (894..1961)
<p>/note="similar to H. pylori 26695 gene HP1365"  /codon_start=1  /transl_table=1  /product="putative TRANSCRIPTIONAL REGULATOR"  /protein_id="AD07036.1"  /db_xref="GI:4156082"  /translation="MOKKIFLDEDDVILSESIEKFEHLGVEFCANFGKAEHRLSV  ERNILLVQVDEMSLELFRKIDPLISFVPIITLQDNATKKNFNGASDYL  KRPFDLEFARKREPNDPRLIEMINIVYHQAIVNKKETLAKTAKOLLEFLEH  KGIISQALENNLMEOATDSTLRITIVLRKLKNCIEFHKGVRNPL"</p>		gene	complement (894..1961)
<p>/note="hjp1444"  complement (894..1961)</p>		gene	complement (894..1961)



QY 215 TCTTTGTTCCGAGTCTCAGTGAAGAGACATCTCTCTGATCTTGGGAAAAACA 274  
 DB 8519 AACGCCCATTCAGGTTTGAAGAAAGAGATGAGTCCGAGTTTGGGAAAGACAGC 8578  
 QY 275 TGTACACAAAGGATATCTCAGCATGCTGTGGCAATGCTTAATGTTTGAAGC 334  
 DB 8579 ATTAAACTTCATATTCCTGATGTTGAGGCTTTATTTAGATGAGGCTTTATGCA 8638  
 QY 335 GTATATTAATGATTGGAGCGCTATGCTTGGAGCTGTTCTTGAATCTTTGCTT 394  
 DB 8639 CTTTATTAATCTCAGTGGCGGATGCTTGCATGCGGTAGTGGTCAATCTTTTGTG 8698  
 QY 395 ATCTGGCAGCTACAGTATTGGATGCGCACCTACCTGTAGAGCTGCGGGGAT 454  
 DB 8699 ATCTGGCGGTCATGCGATTTTGGAGCGAGCTGATTTTACGGGAGTGGCGGAT 8758  
 QY 455 GTTCTGCTATGGGATGGCGTATGATGCAATGTTCTTGTATTCGAAGATCCGAG 514  
 DB 8759 GTTTTAAACGTTGGGATGTCGTGATGCTATATCATATCAAGCAAGCATTAAGAA 8818  
 QY 515 GAATTTTATGTCCTCAAAATCTTAAAAATCTGAGAAAAAGATATCAAGGCTTTT 574  
 DB 8819 GTTTTAAAGAGAGAGAGGAGTGTGTTAAGCGCATCCATTAGGCTATATCAATGCGAGC 8878  
 QY 575 GGAGCCATTTTGAATCTTACTGATCTACATATTTGGCTGAGCACTTTCTTCTCTA 634  
 DB 8879 CGGCGATTTTGAATCTTCAATATCATCTTGTATGCTTGGGTGCTATAGCTTAT 8938  
 QY 635 GATACAGGGCTTATTAAGAGTGTGCTTGAATGATTTAGAAATTTCTCTCAATG 694  
 DB 8939 GGACAGAGAGATTAAGAGCTTGGCTCCTCACACAGGCAATTTGGATTTTGGCTATC 8998  
 QY 695 TTAAAGCTCTTTCATGATTAATTTTCTTCACTGCTGATGATTAAGACCAACAT 754  
 DB 8999 ATCAACGCTATTAATAGGACCAAG-----GGAATTTTCAAGCCTTTTAA 9043  
 QY 755 ACAAGTTGATATGATTAATTAAGTCTGCGGAGTAAAGCATGATTTCTTGAAGATC 814  
 DB 9044 CCTAATTAAGCCCAACAA-----GCTTTACTTTTGGTTGGCTGAAATA 9095  
 QY 815 AAAAACTTTGGGCTGTTTGAAGTGTGTTTCTTTAGGTTGCTGCTCGGGGTTT 874  
 DB 9096 AAAGAGCTTGAAGGTTTATGGA-----TTATTTCAACAAGTTGAATCTTAAGTT 9149  
 QY 875 GGAGCTGGAATCCGTTTGGGAATGATTTTAAAGAGGATATCTTACTTAT 934  
 DB 9150 CATCGTATTCATATTAATTAAGGATGATGTTTCAAGCATCTTGGGCTTCTGGGCTT 9206  
 QY 935 CCAAAAGCATGCGATCAAGATGTTCTCAATGCTGCAAAAGTTGTCATTAACATA 994  
 DB 9307 -----AGGGCTTTTGTTTTAAAGGGTTTCTTTAGGATTTTGGCGGGGAGT 9261  
 QY 995 CAGGAAGCTGCTCTTCTTTAGAGACTTCGATTTCAACATTTGATCTTCAAGAAAG 1054  
 DB 9262 TTGGTACAGGTCGTTTACATCAAAACCCCATTAAGAGTCCGGATCTGTGGA 9321  
 QY 1055 ATCAAAATCTATTTAGTAAAGCTTAAAGTATTAAGTAAAGAGATCAAGCTCTCTC 1114  
 DB 9322 AAAGAGCTGCTTTTAAAGGCGTCAAGTGAAGCAATTTGATCTTAAAGAAATTTTA 9381  
 QY 1115 CTAAATTAATGATCATGAGCTGCGGTATTTGTGGATTTGTTCAGAAACAGGCTTGA 1174  
 DB 9382 ATCAAAATCTCTTTTGTGAAACGCGCAAAACGAAGTTTGAAGCTATGCTGCTAAC 9441  
 QY 1175 TTTCTCTACGAACTCTTAAAGCAAGCAAAATTTTGTCAAGAGTAAAGCAAACTA 1234  
 DB 9442 ATTCTAAACCAAGCGGATTTTGAATCCGTAAGTTTGAACGTTGGCCCTAGAGTG 9501  
 QY 1235 TCGAAGAAATGCTTATCAGCGCAATCGGCTTTTGAAGAGCTTTGGCAATCTTGT 1294  
 DB 9502 GGGAGCGAATTTGAAGAAAGGAGCTTTTGTGCTGATTTTGAAGCTTATATGAGCATATG 9561  
 QY 1295 CTATATGAGTTTGGCTTTGAATGCGAATATGCTTTCAGTCCGATATGCGCTTATAT 1354

DB 9562 GTTATGAGTTCGCGTATGATGAGCGCTTTCGCTTAGCGAGCTTGTGGCTGTG 9621.  
 QY 1355 CATGACCTTTTGGCTACCTGTGAGCTTGTATTATAGCAATTTCTTTGAAGAAAT 1414  
 DB 9622 CATGA-----TGTATTTTATGAGCAAGCTCGGTGATCTTTTAAAGAT 9666  
 QY 1415 CAATAGATTTGAAGCAATGCTGTTTAAAGATGATGAGTATGAGGATTAACAT 1474  
 DB 9667 GATATGAATTTGAAGTATGAGGCTTGTCTACATTAATGAGGATTAATGAT 9726  
 QY 1475 ACTTATCATTTTATGATGATGATTCGTAAGATCGCAAGCAAGCTTTTACCCCTAG 1534  
 DB 9727 ACATCATTAATCTTATGATGATCAAGAAAGATGCTCTCAAAAAACAAAGCGC 9786  
 QY 1535 CATGTTTATTAATGATGCTTCAAAAGAGCTTACCGCGCAAGGTAATGACACAGCT 1594  
 DB 9787 ACTCAAGCATGATGATGATGATTCAGACAGCTCACGCGCAGCTTTTAACTTCTTA 9846  
 QY 1595 ACAATCTATCACTTTTGTATATGCTTTTATAGCGGCTCCCTGCTTATTTT 1654  
 DB 9847 ACCGTGTTTGTGATGATTTGTGCGTGTGGAGTAAATCATTAATGCTT 9906  
 QY 1655 GCATTTATTAATGACCAATGAGGATTTCTTGAAGCTTATGCTCTTATATATGACCA 1714  
 DB 9907 TCATGCGCATGCTATATGACAGATTTGAGGACTTATAGCTTATTTTCAATGCGCCCT 9966  
 QY 1715 CTTCTGTTGTTTATGATGCTGTA 1740  
 DB 9967 AAAGTAGCTTATGTTAGGCTTGA 9992

RESULT 8  
 CUI1168X4/C  
 LOCUS CUI1168X4 282183 bp DNA linear BCT 09-JUL-2002  
 DEFINITION Campylobacter jejuni subsp. jejuni NCTC 11168 complete genome;  
 segment 4/6.  
 ACCESSION AL139077 AL111168  
 VERSION AL139077.2 GI:6968444  
 KEYWORDS  
 SOURCE Campylobacter jejuni subsp. jejuni NCTC 11168.  
 ORGANISM Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 Campylobacter.  
 REFERENCE  
 AUTHORS 1 Parkhill, J., Wren, B. W., Mungall, K., Kelsey, J. M., Churcher, C.,  
 Basham, D., Chillingworth, T., Davies, R. M., Feltham, T., Holtroyd, S.,  
 Jagers, K., Karlyshev, A., Moule, S., Pallen, M. J., Penn, C. W.,  
 Quail, M., Rajandream, M. A., Rutherford, K. M., VanVleet, A.,  
 Whitehead, S. and Barrall, B. G.  
 TITLE The genome sequence of the food-borne pathogen Campylobacter jejuni  
 reveals hypervariable sequences  
 JOURNAL Nature 403 (6770), 665-668 (2000)  
 MEDLINE 20150912  
 PUBMED 10688204  
 REFERENCE 2 (bases 1 to 282183)  
 AUTHORS Parkhill, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-FEB-2000) Submitted on behalf of the Campylobacter  
 sequencing team, Sanger Centre, Wellcome Trust Genome Campus,  
 Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk  
 COMMENT Notes:  
 Details of C. jejuni sequencing at the Sanger Centre are available  
 on the World Wide Web.  
 (URL, [http://www.sanger.ac.uk/Projects/C\\_jejuni/](http://www.sanger.ac.uk/Projects/C_jejuni/)).  
 FEATURES  
 source  
 1 282183  
 /organism="Campylobacter jejuni subsp. jejuni NCTC 11168"  
 /strain="NCTC 11168"  
 /db\_xref="taxon:192222"  
 complement(43..1443)  
 /gene="Cj1009c"  
 complement(43..1443)  
 CDS

```

/gene="Cj1009c"
/notes="Cj1009c, unknown, len: 466 aa; 29.1% identity to
HP0282"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj1009c"
/protein_id="CAB73265.1"
/db_xref="GI:6968445"
/db_xref="SPTREMBL:Q9PNT1"
/translacion="MNNILIIIDGILAKHFLERLCFEKGLGYFFTVVCQSEKNKLNLI
SSEYIDHYEDPTSTARLENIMSKDFQAFPMODEFETKSKYEALRSNPNLEIEM
DFWGLSVNDTHANLADARMTLSREFMDFLDIALTAOYIGLGVGEINSEVKIPAGSIFA
YRHLSSQQRWRVILYRNSKIYFVPSFVLENDLILVDFVILQSIFNIRGKA
QDPMPGSGNFADIDMKNQNNQDERVLDITLKLTQKSNKARFFIHIIVPKLGVWYE
KLKLSDEKGVDFNTDFQISMMQNNIDGLVTDIKNFEKERQAFDLKIPITM
KVGASDELKEAILLSADESELENANVITDLSKOLDFGVILYYNPNSONTDMEE
YFRSLKLYDKNIQIINKNDENPLNLQYREDLLQFVSFQKELLNRDFARNLSTNLR
HYKMRONYOLFVPE"
/1473..2594
/gene="tgt"
/1473..2594
/gene="tgt"
/EC number="2.4.2.29"
/notes="Cj1010, tgt, probable guanine
tRNA-ribosyltransferase, len: 373 aa; similar to many e.g.
TGT_ECOLI guanine tRNA-ribosyltransferase (EC 2.4.2.29)
(375 aa), fasta scores; opt: 1016 z-score: 1190.9 E(): 0,
41.3% identity in 373 aa overlap. 56.7% identity to
HP0281"
/codon_start=1
/transl_table=11
/product="guanine tRNA-ribosyltransferase"
/protein_id="CAB73266.1"
/db_xref="GI:6968446"
/db_xref="SWISS-PROT:Q9PNT0"
/translacion="MEFKLHKQGMARVCBITTAHSTFLTPVPMVGTGAVKSLDAN
DMKNELDAKILANTYHMLKRLKPSKVDQFGLHGFTRKDRSFITDGGQAFSLSN
SKHNEGIEFKSHIDGSRHLFTPKSVLDITQYDENSIDIMLIDLVALPATKERVKISV
DITLWAKETIYTHKNQKIGIQINIFGI IQGTDYERKRCALSLNMPDFGLAI
GGLSGRENALMYETQNLNPDENRPRVLMGVGTPEDLVENVERGVDMFDCVMPTR
NARGTPTFTSGFKNIKKAFINDHVIDSTCSYCTCRNPSRGVNLHFKAKELTPFR
LASLHNHYLELAKRREAILNNSFTQFRNFVHLRGK"
/2596..3363
/gene="Cj1011"
/2596..3363
/gene="Cj1011"
/notes="Cj1011, probable membrane protein, len: 255 aa; no
Hp match"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB73267.1"
/db_xref="GI:6968447"
/db_xref="SPTREMBL:Q9PNS9"
/translacion="MLSLHNLKILQAKNLETYSIYQKIFVYVGLNLEWLFND
EKIYKQDGFSEIEEFNFIKSDLEFKVQNTHFQNLHEKIGILKGGFKNFY
KKSFLVKQKINKNLQNLSSAFNLLSSQAOYKHLKILNLSILSKNTKEHLAR
IDTLYTNAIKNEKNKNSIVLLSSIFLPLNLIVGFGMTNNLFFKXDSPVGTLY
IFSLICLILVGFIFYYSKTKFEDLDEGKAKQTK"
/3416..3432
/gene="tRNA-Arg"
/3416..3432
/gene="tRNA-Arg"
/product="tRNA-Arg"
/notes="tRNA Arg anticodon CGG, Cove score 83.29"
/complement(3495..3944)
/gene="Cj1012c"
/complement(3495..3944)
/gene="Cj1012c"
/notes="Cj1012c, possible membrane protein, len: 149 aa; no
Hp match"
/codon_start=1
/transl_table=11

```

```

/product="putative membrane protein"
/protein_id="CAB73268.1"
/db_xref="GI:6968448"
/db_xref="SPTREMBL:Q9PNS8"
/translacion="MVIFCAALALLPLGVIAVLITSDGKTTIKKOKTSOKOVAEKT
KFEPTDLDMIIAASDVKLTDIELKELAKLYVOTHLGSKTSKELDAEAAKKLEFVSA
LAANINASQTVSYLNKELKXIYSYKKEIDAYEHMGLAKRKIKEDK"
/complement(3961..7206)
/gene="Cj1013c"
/complement(3961..7206)
/notes="Cj1013c"
/notes="Cj1013c, probable membrane protein, len: 1081 aa;
contains three domains; aa 1-90 contains three membrane
spanning domains; aa 90-780 is non-membrane, and aa
780-1081 contains ten possible membrane spanning domains.
Some similarity in C-terminal membrane domain to eukaryotic
proteins e.g. CCSA, CHLRE cytochrome C biogenesis protein
CCSA (353 aa), fasta scores; opt: 459 z-score: 515.4 E():
2.1e-21, 36.0% identity in 267 aa overlap, and to
prokaryotic equivalents e.g. CCMF_ECOLI cytochrome C-type
biogenesis protein CCMF (647 aa); BLASTP scores; E = 0.20,
24% identity in 293 aa overlap. 39.0% identity to HP0378
(domains 2+3 only)"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB73269.1"
/db_xref="GI:6968449"
/db_xref="SPTREMBL:Q9PNS7"
/translacion="MKNIIKSGDLRVSVFLFALFCALATFIESAYGTPTWAMV
YDTWFYEQLLGILNLCGMFRIKMGFLKLPIMPHISFLFILGVSAMTRYAGFEG
ILPIREHTQNSLIESKTSLRISAIDGERYSANDRYIGNLPFANSFKLNLGDDQ
AVLKQDLILNAHYTKENNSDPLVLDSQSGVDVKFEKVEIGNVFAFM
NDNVKAPFVKIDENLTLSSSENHLFLSMLDQNLDLKIGKANAKERRLYEINDISFV
VKAASLHAQALEGSRNPODESFLHFKSAWLSVGRMTLJSTGEPQNMKNSLLHKK
DFALSNEKLELTGSNALKLSYKNESKEFTFYFNKFMIELADQKFEISWALSY
QEPFDIYLRDFVLDYRPGSMSPAYSEITVKNNENFDYRIPMNVLDYDGVRFYQ
SSVDQDEKTVLSDKDPKPTIYIGVFLCLGMFNFLNPHSRFTLARLNKDTLK
HTSVITFILLSSFGSEKTAQDLNSTLPPVNTNHAALATLIVQKSDAGRMVPTDLS
REILEKHOSDSYKQNSNAVMSLMDVDKQWLEPEFLMPQNAVRDAIANILEIPS
AKIAYKQDFDENRYKLOKYVENAKRNPARGVDFKEILKUDERANVNLVPSGL
FKFIPQNNPNVLAFFSAVTTLKDGEGHIVALIQNYFSAVENAFQGNWIRADSG
LKFIKEYQEKIGYKMPKSTKMEIEIFSNAKEIFVKLAPVYLIAGFLLILVFSKMVY
PNLKSIFKVVYLVNLAFFVITVGLGRAYLSHAPNSNGYESMYIYIAMSLSGI
FFSRKSPIALSLSILSGVLMVAHLSMMNPQITNLVPVLSYMLSIHVSIVTASGYF
LGI CALLGIPTFLMCLFKDQKYNILNRLNITEATRINEMAMTFGICLLTVGNFLCA
IWANESGRYMSHDSKETHALVSLIYAAIHLHRLMPKCYCNQFVALMSFAYVITM
TYFGVNYFLGLHSYAGAAQIPNYVYMGFALMVVLAFFARRKRNFGVL"
/complement(7224..7919)
/gene="livF"
/gene="livF"
/notes="livF"
/notes="Cj1014c, livF, probable branched-chain amino-acid
ABC transport system ATP-binding protein, len: 231 aa;
similar to e.g. LIVF_ECOLI high-affinity branched-chain
amino acid transport ATP-binding protein (237 aa), fasta
scores; opt: 686 z-score: 797.6 E(): 0, 46.4% identity in
233 aa overlap. No Hp ortholog. Contains PS00017
ATP/GTP-binding site motif A (p-loop), PS00211 ABC
transporters family signature, and Pfam match to entry
PF00005 ABC tran"
/codon_start=1
/transl_table=11
/product="branched-chain amino-acid ABC transport system
ATP-binding protein"
/protein_id="CAB73270.1"
/db_xref="GI:6968450"
/db_xref="SPTREMBL:Q9PNS6"
/translacion="MLIVKDLHVYVGLIEAVKIGDPKVTGHIVSLIGSAGKSTSL
NALLNSVKRTGEVNFGLYDTRKHLTHTLVOKGIALVPEGRVFINLSVEENLKICAFN
NAENYBHLREQMYKLPFRASKKLAGLTSGGGAQMLAISRALMSFKLLMLDPEPSL
GLAPKIVGEVDFIIVRLKEEGITILLVEQNAYSALKISDYAVYLENGRIVMENBAKNL
IGDDEIRKKYVGL"

```



```

misc_feature      complement (7299..7841)
                  /gene="livF"
                  /note="Pfam match to entry PF00005 ABC transp. ABC
                  transporters, score 147.40, E-value 2.5e-40"
misc_feature      complement (7482..7526)
                  /gene="livF"
                  /note="PF00211 ABC transporters family signature"
misc_feature      complement (7797..7820)
                  /gene="livF"
                  /note="PE00017 ATP/GTP-binding site motif A (P-loop)"
gene              complement (7906..8676)
                  /gene="livG"
                  /note="Cj1015c, livG, probable branched-chain amino-acid
                  CDS
                  /note="Cj1015c, livG, probable branched-chain amino-acid

Query Match      8.4%; Score 148.2; DB 1; Length 282183;
Best Local Similarity 46.1%; Pred. No. 8.2e-23;
Matches 722; Conservative 0; Mismatches 778; Indels 67; Gaps 4;

Qy 161 TTTACCAACCGTGAAGTGAACAACCTGCGCTCAATTTAAATCTGAGCGATGCTTTT 220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86752 TTTACTCAAGAAAGACGCGATGCGCGTACCTTTAAGAGTGGAGCTTTCCTGCG 86693
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 221 GTTCCCGAGTTCACAGTGAAGACATCTCTTGATCTTGGGAAAAACAATGTACA 280
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86692 CCTGTAAATTAATCTTGAACAAAGACATGACCATCTTTAGGCTGATATTTAA 86633
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 281 CAAGGCAATATCTCAGCATGCTGCGCTTGGCAATGCTTATGTTTGGATGAGCGTAT 340
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86632 ATGAGTATGATAGCTTAAACGGTGATCACTTTTATATGATATTTATATGATGAT 86573
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 341 TATAGATTTGAGGCGGTCAATGCTTGGAGCTGTTCTTGAATCTTTGCTTATCTGG 400
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86572 TACGCTGAGGAGAAATTTTGCAAAATTTGCAATGCTAGTTAAATGTTTGATGAT 86513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 401 GCACCTCACAGTATTTGGAATGCGGCACCTGTCAGGACGCGGAGATTTGTTCTT 460
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86512 GCCGTATAGCGTATGTTTGAAGCACTTTAACTTTACAGGAGATGGACCTGTTTAA 86453
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 461 GCTATGGGAGTGCCTGATGATGCAAAATGTTCTTGATTCGAAGATCCGAGAGAAATT 520
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86452 ACTGTGGGTATGCGATGATGCAAAATGATATATCATGAGCGATATCGAAGACTCTG 86393
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 521 TTATTTCTCAAAGCTTTAAAAATCTGTGAAGAAAGATATACCAGGCTTTTGGAGCC 580
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86392 CGAATGAGATTAATATATAGTTAGTATAGAACAGGTTATTAATAAATGCTATGATG 86333
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 581 ATTTTGAATTTCTAATCTGAATACAGTATTTGGCTCAGCACTTTTCTTCTTGATACA 640
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86332 ATTTATGATTTCAAAATATCACTTTTATGTAATCTTGGAGCGCTTATGCTTATGAA 86273
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 641 GGGCTATTTAAAGGTTGCTTTGACATTTTGAATTTTGAAGATTTTCTCTCAATGTTTAC 700
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86272 GGAGCAGTTAAAGTTTGTCTGTGACTTTGGAAATGGTATCGTTGTTCTATGATTTACA 86213
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 701 GCTCTTTTCATGACTAAATTTTCTTCATGCTGTGATGAATTAAGACCAACATACACAG 760
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86212 GCTATTTTAAAGAACTCATGAAATGTTGATTTTATTAACAAACGATAGAAAAATATAT 86153
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 761 TTGCATATGATGATTAAGTTCGTGGGATTAAGCATATTTCTTGAGAGATGCAAAAAA 820
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86152 AATACAAATTTTGGTTGTTATAGAGAAATATATACAGTTTATTTAGCAAAAAA 86093
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 821 CTTTGGGCTGTTTTCGAAGTGTGTTTCTTTTAAAGTTCGTTGCTCTCGGTTTGGAGCC 880
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86092 TTTATGATTTTATGAAATGCGTTTGTCTGATTTCCCTTTC----- 86050
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 881 TGAATTCGTTTGGGAATGGAATTTTAAAGAGGATATGCTTTACCTTTAATCCAAA 940
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 86049 -----TTTATTTTGTGTTTGTGTTGTTATTTATTTACTTTGG 86013
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 941 GAGCATGCAATCAGGATGTTGCTCAAAATGCTGGCAAAAGTTTGTGATTAACATACAGAA 1000
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 86012 GATAGGGGTTTGCAATATGCAATGATTTAGCGGAGGAACCTTAGTCACTTAATAT 85953
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1001 GCTGTCTTTCTTTAGAGACTTCCTGATTTCAAACTTTGGATCTTCAGAAAAATCAAA 1060
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85952 GAAAAATGACAGCTCCTTATTAACGCAAAATTCGAGAAATTTTGAANAATCAAGAACTTTTCA 85893
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1061 ATCTATTTTATGATTAAGCTTTAAGCTTATACATACAGATAGACCTCTCTCTTAA 1120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85892 AATTTATCTGTAACGTAGTTTGAAGCAAT--GAAAGAACTATACGCTTTTATAGAA 85835
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1121 TTAACGATCAATGAGCTGCGCTTATTTGAGGATTTGTTGTCAGAAACAGGCTTATTTCTC 1180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85834 GTATATGATATGTAAGCAATGATATAGTATAGATATTAAGCACTC-----TTTAA 85785
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1181 TACGAAACTCTTAACGAAACGCAAAATTTTGTCAAAGTAAAGCAACATATGAGAG 1240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85784 AAGGATACGGGAAAAATTTGAAGTGCCTGCTGTGATGATGATGCTCAAAAGTATGAT 85725
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1241 AAAATGCTTATACAGCGACCATCGGCGTTTGAAGCTTTGGCAATCATCTGCTAT 1300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85724 GAACCTTAAATTAAGACTTATAGCTATAGCTGATCTTGAATAGCAATTTTATAT 85665
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1301 GTAGTTTGGCTTTGATGGAATGCAATATGCTTTCAGTCCGATAGCGCTTATATCAATGAC 1360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85664 ATAGCGTTGGTTTATGATGCGCTTTTGTCTTCCGCTATATCATGTAATTCATGAT 85605
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1361 CTTTGGCTACCTGTGAGCTTGTGTTTATAGACATTTCTTTTGAAGAAATTCAAATA 1420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85604 GTTGTGATTTAC-----TTTGAAGCAATTTCTTATTTTAAATTTGATGTC 85560
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1421 GATTTGCAACCAATGAGCTTATATGATGCTGATTTGGGGATTCATTAACAATACCTTG 1480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85559 AATCTAGATACCTTGTGCTGCTTTTAACTGCTGTGGAATCTTTGAATGATACAT 85500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1481 ATCATTTTGTATGCTATTCGTGAAGATCGCAAGCAACCTGTTTACCCTATGCAATGT 1540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85499 ATATATTTTATATGAATCAAGAAAGTATTAATAACAAGTAAACCAACCTTGACCT 85440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1541 TTAGTATATAGTCCCTTCAAAAGACGTTACGCCGCAAGCTTATGACACACTACACT 1600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85439 ATTTATATGAATAAGTGTTCAGCAACTTTTATCAAGAACGTTTAACTTCAGGGTGACT 85380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1601 CTATCACTTTTGTAAAGCTTTGTTATATAGGGGCTCCCTGTCTTATTTTGAATTT 1660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85379 TTAGCAACCGTGTGATACCTTATTTCTTTGGGGGAAATGATACAGATTTTCTTG 85320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1661 ATTTAGCAATAGGATTTCTTACGAACTTTATGCTCTTTATATTTGACACACCTCTG 1720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85319 GCTTTATTTGATGATTTATAGGGGAACTTTAAGCTCTATTTTGTGCAAGTCCGACA 85260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1721 TTGTGT 1727
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85259 TTATAT 85253
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 9
AE000652/c 15058 bp DNA linear BCT 06-APR-1999
LOCUS Helicobacter pylori 26695 section 130 of 134 of the complete
DEFINITION genome.
ACCESSION AE000652 AE000511
VERSION AE000652.1 GI:2314720
KEYWORDS Helicobacter pylori 26695.
SOURCE Helicobacter pylori 26695.
ORGANISM Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
          Helicobacter.
REFERENCE 1 (bases 1 to 15058)
Tomb,J.F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
Fleischmann,R.D., Ketchum,K.A., Klein,H.P., Gill,S.,
Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,

```



Glodek, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Uterback, T.R., Peterson, J.D., Kelley, J.M., Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C.

The complete genome sequence of the gastric pathogen *Helicobacter pylori*

Nature 388 (6642), 539-547 (1997)

97394467

9252185

2 (bases 1 to 15058)

Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G., Glodek, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Uterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Weidman, J.M., Fujii, C., Bowman, C., Matthey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C.

Direct Submission

Submitted (08-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

3 (bases 1 to 15058)

White, O.

Direct Submission

Submitted (17-MAR-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers

1. 15058

/organism="Helicobacter pylori 26695"

/strain="26695"

/db\_xref="taxon:85962"

complement (111.968)

/gene="HP1538"

/note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:J03176 PID:152084 SP:P51131 percent identity: 29.8%; identified by sequence similarity; putative"

complement (965.2203)

/gene="HP1539"

complement (965.2203)

/gene="HP1539"

/note="similar to GB:J03176 PID:152084 SP:P51131 percent identity: 39.09; identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=1

/product="ubiquinol cytochrome c oxidoreductase, cytochrome b subunit (fbch)"

/protein\_id="AAD08579.1"

/db\_xref="GI:2314721"

/translation="MASIKKAKNLGELMDRLGNKLVKVLMTYWIPIKNIFLWANG VILLTGLVGVGIFLLMYKPKAKAFDSVNETIQEYVAYGCHLWHEMHTAASMP VLIYHMFVGIYSGYKGRMINISCHILFVFSAAFGSYMLPWQMSWAAAVIT NLFQGIPIGADVVEVIRGNVADSTVFHMLHVFLLPIAIIILLVGVFYSIRIPH VNNQGEIEDEBEKEFIEGKESKVIPEWPFVLSKDIFFVCAFVFFYLVYHY DFANDPINFERSANSLKTPPHIYPEWFLWSVEVLRGFFFSADLGLMAGFAVQVIFFL PFLDRSPVAPAKRPAFMVFWLVIIDMIVLITVYGLKPLPGIGIKYGLAGSITFLL FFVLPIITAEKSKQGV"

complement (2214.2717)

/gene="HP1540"

complement (2214.2717)

/gene="HP1540"

/note="similar to GB:J03176 PID:152084 SP:P51131 percent identity: 39.22; identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=1

/product="ubiquinol cytochrome c oxidoreductase, Rieske 2Fe-2S subunit (fbcf)"

/protein\_id="AAD08580.1"

/db\_xref="GI:2314722"

/translation="MADIQRDRFLGMSLAVTAIGAIASLVAMKKTWDPVSVSAGF

TTTIDVAMQEQGSTVEWROKPYVILKRSKKEGFNEKRDPKVGESVPTTATQIOTHUG  
CITYQDEEKGFLPCPCGSRFTSDGVNIAGTTPPPRPDPDIPPKIEGTKITFGEAGAY  
KQWAKA"

complement (2842.5841)

/gene="HP1541"

complement (2842.5841)

/gene="HP1541"

/note="similar to GB:L42023 SP:P45128 PID:1007188  
PID:1221387 PID:1205499 percent identity: 37.66;  
identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=1

/product="transcription-repair coupling factor (trcF)"

/protein\_id="AAD08581.1"

/db\_xref="GI:2314723"

/translation="MIQSYRALNKGFDYOILACKDPKESELAKEVISYFPKNTKAI  
LFPEFRKNDLRSFPEFLQGLREFVQALENKQETIITAPIALHLPKPKEL  
LESFKITLLKYNLKKDLKFLYGYGIELDVEGEASFRGDIVDIYAPNSKAYRLS  
FDTCESEIKFDPITOMSLKDELLEIPIPTLPSLDESSYKDLTKVEQSPLENFSK  
DLTSPGLWFLGKAQDILLIYKSIISPRALKEEIQELASNELDCERCFKVLVNAQK  
YEDLETHALEGFIALHSHKITALNPNKTIIDNAISALDAGNEMCVIAPFVLENFT  
PDGIFITLSNFSFERKKRKSKALNELNPGEWVVDYGVGVSQVQHSVLGSKRDF  
LETAYLGEDKLLPVENLHLIARYVAQSDSVPAKORLKGSPFLKAKVTRTKLEIAS  
KIIEAERNLILKQMDVHLEAVFVSHAGFEYTSQDKATAEISKDSHRVMDR  
LLSGDVGFGTEVAMHAIFCAFLNGFQSALVPTLLAHQHFETLRARFENFGVXVAR  
LDYASEKKNLLKAVELQVDALIGTHAIGAKFNGLVVDVEEHFVGKQKALKE  
LSKSVFLSMSATPIPTLNALSOIKGISLKTPTDRKPSRTFLKEKNDLLEKEII  
YRELRRNGOIFYTHNHASILKVKTKLEDILPKLIAILHSHQINANESSEITLFFAKG  
NYOVLICTSIVESGHIPLNANTIIIDNAQFGLADLHOLGRVGRGKGEFCGYFLIED  
QKSLNQALARKLALKEKSNYLGSGESVAYHDLIRGGGNLQDQSGHIGFVLYED  
RMLEDAIYELSGKKRLKSEVEIQVSAFNPELIASLRLDLRYRLSCENTDEV  
QIHEEIERDFGKIDLSAQFLQIITLKILANQLGIKLSNPNQNTITYSDEKKSLE  
KAPSKDDNDILETLKHLRAQISLKRR"

complement (5845.6255)

/gene="HP1542"

complement (5845.6255)

/gene="HP1542"

/note="hypothetical protein; identified by GeneMark;  
putative"

/codon\_start=1

/transl\_table=1

/product="H. pylori predicted coding region HP1542"

/protein\_id="AAD08589.1"

/db\_xref="GI:2314731"

/translation="MAIFDNNKNSAKAKTGPATIIAGTKIKGELHLDYHLHVDGELE  
GVVHSTVTVIGOTSGVGEIETNKLVLVSGKFTGVEAEVVEIMPLGHLDGKISSQEL  
VVERKGLIGETPRKNIQGGALLINEQEKKIENK"

complement (6174.7112)

/gene="HP1543"

complement (6174.7112)

/gene="HP1543"

/note="similar to GB:U07173 PID:460955 PID:1100877 percent  
identity: 37.15; identified by sequence similarity;  
putative"

/codon\_start=1

/transl\_table=1

/product="toxR-activated gene (tagE)"

/protein\_id="AAD08582.1"

/db\_xref="GI:2314724"

/translation="MFLDRRLIYMTVDSKGSRYINVHLFRQIGLYALLSVGSLFLF  
GISLLVNLQEKIDKQHALITKEFEKKETNEKLSLQMDFDLDDQLSGERINDLEE  
VGVNRPPEERKEGNSRRLDVAGITGLQKSFIMRLIPNDYPLESVGAAPNKRTH  
PIHLVHNHTGLDLSTAINTPVASAGVGLASKGWNGYGNLIVKVFHPGKTYVA  
HLNKIVYKTCGEFYVKGQQLIGYSGNTGMSFPHLYEVRFLDQPINPMSFTNNMKDPE  
EVFNKRSRWQSLITIINKLMQDQRLSSLKAQK"

complement (7121.8047)

/gene="HP1544"

complement (7121.8047)

/gene="HP1544"

/note="similar to GB:U07173 PID:460955 PID:1100877 percent  
identity: 31.25; identified by sequence similarity;  
putative"

/codon\_start=1

```

/translate=11
/product="toxR-activated gene (tagB)"
/protein_id="ABD08583.1"
/db_xref="GI:2314725"
/translation="MPQNLVITIDSGSKQKFSKLNKLNLIISVILFLLVGLV
GFLKPLAKMDMTAEARNALRDRPDLYQKNYLTKEIKKREELFVQKIKTISL
IEVRKANGCVHLVDEVDLNLAKHLLMLIPGMPIKTSALRPTERRHPKIK
IKGVESGILFIAPLNPVYASADGIYDFVATNSNVGKLVRIEHAQPSSTIYTHD
VNVQPSFIOKGQILIGYSKSGNSGGKELFVVRFGKILIDACKFLAMDHPQSALE
ENKFIKWLFWYLEDIVQLQEHVDKALISQ"
/length=8047
/genes="HE1545"
/complement(8047..9231)
/feature="similar to GB:L42023 SP:P43775 PID:1007194
PID:121390 PID:1205502 percent identity: 35.17;
identified by sequence similarity; putative"
/codon_start=1
/translate=11
/product="Folylpolyglutamate synthase (folC)"
/protein_id="ABD08584.1"
/db_xref="GI:2314726"
/translation="MKNSPLNGLNGLAKFLETKPKYHKFPERFIOIYDFQNAFEE
IOAKVHVGTNGKSGTGRFLTLIDONKVLHPTSPPHVEFEREPFLNGSVGSV
LENAHQOLSHAFSSACSVEYATILAVMLAKODVILVAGIGEDSTNALKTIS
VTFPIDHKEFLGDSLESIQTKLAKMSGLSTIAQOEVLVNAOKIAKEXAKLIV
VQNEISKVRDIYERHILARFLAMNEVALKAEFTLLPCNKQVLLNKLKPLIGCE
LSPNLIIDVGNPHSAKALKEIKRIFNKIILILYNYCYDKDAFLVLEILKLVIKV
LILEHEHVIKLEKLGILETGLFEVALFEDVEENYLVGSPFLVANAFLRYQYEK
RD"
/length=9221
/genes="HP1546"
/complement(9221..9733)
/feature="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/translate=11
/product="H. pylori predicted coding region HP1546"

Query Match      7.9%; Score 146.4; DB 1; Length 15058;
Best Local Similarity 53.7%; Pred. No. 2.8e-22;
Matches 303; Conservative 0; Mismatches 261; Indels 0; Gaps 0;

QY 155 GGAATAATTTACCAACCGTAGAGAGCAAACTGCTCAGATTAAATCTGAGCCATG 214
      |||||
DB 14132 GGAATATTTACGCGGCTCAGCGAGCGATTAGCATGCTTTAAGAGTGGGGCGATG 14073

QY 215 TCTTTTGTCCGAGTTCACAGTGAAGACGATCTCTTGATCTTGGGAAAAACAA 274
      |||||
DB 14072 AGCCCTCCCATTCAGGTTTGAAGAAAAAGATTATAGGCCCAAGTTTAGGAAAAAGACAGC 14013

QY 275 TGTCACAAAGCATTAATCTAGCATGCTGTCGCTTGGCAATGCTTATGTTTATGAGC 334
      |||||
DB 14012 GTTAAACCTTCATTAATGCTCTAGCTTGGGGCTTTATTTAGCATGGCTTTATGCTG 13953

QY 335 GTATATATATAGATTGAG3CGTCATCGCTTGGGAGCTGTTCTTGAATCTTTTGCT 394
      |||||
DB 13952 CTTTATTACTATGCGGCG36GTATCGCTTGTGGCTTAGTGTCATCTTTTGTG 13893

QY 395 ATCTGGGAGCTCTACAGTATTTGGATGGCCACTCAGCTTGTGAGCACTGGCTGGGATT 454
      |||||
DB 13892 ATTGTGGCGTCATGCGCAITTTTGGAGCGACGCTGACTTTAACGGGAAATGGCGGATT 13833

QY 455 GTTCTGTGCTATGGGAGTGGCCGTAGATGCAATGTTCTTGATTGAAAGATCCGAGG 514
      |||||
DB 13832 GTTTTAAACCGTGGGATTTGCCGTGATCTTAATCATCATCAACGAGCGCATTTAGAGA 13773

QY 515 GAATTTTATTTGCTCAAAAGTCTTAAAAATCTGTAGAAAAAGATATATCAAGGCTTTT 574
      |||||
DB 13772 GTCTTAAGAGAGATGAGGCGATGCTTAAAGCATTCATTTAGGCTATATCAATGCGAGC 13713

QY 575 GGAGCCATTTTGTATTAACCTGACGATGATTTGGCTCAGACATCTTTTCTCTCTA 634

```

```

DB 13712 CGGGCATTTTGTATTTAATATCACTTCTTGATGCTTCAAGTATATACGTTAT 13653
QY 635 GATACAGGCGCTATTAAAGGTTTGTGATCATGATTATTAAGAAATTTCTTTCATG 694
DB 13652 GGCACAGAGAGCGATTAAAGGCTTTGCCCTTAACACAGCATTTGGATTTAGGCTTAT 13593
QY 695 TTACGCTCTTTTCACTGACTAA 718
DB 13592 ATCACCGCTATTGTTGGACGCA 13569

```

```

RESULT 10
AP003192/c
LOCUS      301550 bp      DNA      linear      BCT 10-JUL-2002
DEFINITION Clostridium perfringens str. 13 DNA, complete genome, section 8/10.
ACCESSION AP003192
VERSION    AP003192.2 GI:18146740
KEYWORDS
SOURCE
ORGANISM   Clostridium perfringens str. 13 (strain:13) DNA.
            Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
            Clostridium.

```

```

REFERENCE
AUTHORS    Shimizu,T., Ohtani,K., Hirakawa,H., Ohshima,K., Yamashita,A.,
            Shiba,T., Ogasawara,N., Hattori,M., Kuhara,S. and Hayashi,H.
            Complete genome sequence of Clostridium perfringens, an anaerobic
            flesh-eater
            Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)
JOURNAL     21664373
MEDLINE     11792842
REFERENCE   2 (bases 1 to 301550)
AUTHORS    Shimizu,T.
TITLE       Direct Submission
JOURNAL

```

```

COMMENT
FEATURES
source

```

```

gene
/organism="Clostridium perfringens str. 13"
/strain="13"
/db_xref="taxon:195102"
/note="anaerobic pathogen for gas gangrene"
complement(24..413)
/genes="CPE1825"
complement(24..413)
/genes="CPE1825"

```

```

/note="129 aa, similar to sp:YOHY BACSU HYPOTHETICAL 14.7
KDA PROTEIN IN ACCO-FOLD INNERGENIC REGION from Bacillus
subtilis (135 aa); 40.7% identity in 113 aa overlap. Also
similar to p1r:JC2527 alkaline shock protein from
Staphylococcus aureus. 2 putative transmembrane regions
were found by PSORT."
/codon_start=1
/translate=11

```

```

/product="probable alkaline-shock protein"
/protein_id="BAB81531.1"
/db_xref="GI:18145489"
/translation="NMENMRDANIGYINISDEVIGVAGIASRIDGILEINHNST
GIGHKFSKSLGKIKVNESEAYIEGVYQGIKIPDVVSQVQENRIVEAITG
LKVAVLVIVQNIILIKEDKNEILKK"
complement(498..995)
/genes="sp0111AH"
complement(498..995)
/genes="sp0111AH"

```

```

/note="165 aa, similar to sp:S34H BACSU STAGE III
SPORULATION PROTEIN AH from Bacillus subtilis (218 aa);
22.4% identity in 143 aa overlap. Putative N-terminal
signal sequence was found by PSORT
CPE1826"

```

```

/codon_start=1
/transl_table=11
/product="stage III sporulation protein AH"
/protein_id="BAB81532.1"
/db_xref="GI:18145490"
/translation="MNRKQAGIILTLALIVCTGLATRVNNOIKESMGVPTAFGEN
DDNATERSSNYFESNRNRQKSKTTIDNLKAIVEDKNTSAEQKEAEKELTEKTWAR
DYETRIELSLKSGIEDYCFIDGDNKRVVVKINEELTQEKWVEIQDINVNNSKVYDV
DIEKK"
gene
CDS
complement(1021..1605)
/gene="spolIIIAG"
complement(1021..1605)
/gene="spolIIIAG"
/notes="194 aa, similar to pir:B69712 mutants block
sporulation after engulfment spoIIAG from Bacillus
subtilis (229 aa); 28% identity in 157 aa overlap.
Putative N-terminal signal sequence was found by PSORT
CPE1827"
/codon_start=1
/transl_table=11
/product="stage III sporulation protein AG"
/protein_id="BAB81533.1"
/db_xref="GI:18145491"
/translation="MSKILKQKITNLIILLVIMFVLVSVFTGVNNITKSEKTNL
EKYSKEMDNSQKDSVLVSQYQEKDLEKIGKINGVSDVDVINFQSEKVPVAV
DNSQKSTTBETSEGGTRVNSQETDGDKIWNSSNGSEFVLKTEKPEVLGVWVVA
EGAPDSKYEITKAISSLNYSISVDKYNLAMKK"
gene
CDS
complement(1641..2033)
/gene="CPE1828"
complement(1641..2033)
/gene="CPE1828"
/notes="130 aa, no significant homology"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB81534.1"
/db_xref="GI:18145492"
/translation="MKNKDITVQEKISDVTRSDEKKSLYDEEAILKSVENKLEKS
LKDFEYNEFEVNLKIDFIDVFKFIEKAEITVLDEKVKVDKVVVGKVKVNBKEK
KDSFLKVEKFEKELISEHNIIVSYA"
gene
CDS
complement(2207..3379)
/gene="spolIIAE"
complement(2207..3379)
/gene="spolIIAE"
/notes="390 aa, similar to gpu:AP001516 222 mutants block
sporulation after engulfment from Bacillus halodurans (397
aa); 26.4% identity in 333 aa overlap. Putative N-terminal
signal sequence and 9 putative transmembrane regions were
found by PSORT.
CPE1829"
/codon_start=1
/transl_table=11
/product="stage III sporulation protein AE"
/protein_id="BAB81535.1"
/db_xref="GI:18145493"
/translation="MKKLIINILLSLIFVFCFLKVLANDNLSSKEKIDNDSRIERF
YDINNLETLLGNNSAKEMYINYLESGEDPITLKIGASILSYIFRELNTVLKPI
ASLIVALLSALLKNOADAFIEGVQIAFAFCYALLIMLTSPFISINLGEVIT
SIIDFMNVIYVIMLJATSGVTSATIDPILGAVSITPRIYTFELFLIAYFTL
QFVNLSERKIDRMCKFIQVNMISQGFILITFVGLTURGATDLDVAVATKVF
AVDNFVIGKAFSDATIDVAGVSLAMKSVITSLGVILVVIYVPIIKIGLMAICFK
LTSAPIEPVDSKISSVETVGEALFMINSCTISVSIMFFIMASMMATAGNFIVGG"
gene
CDS
complement(3390..3776)
/gene="spolIIAD"
complement(3390..3776)
/gene="spolIIAD"
/notes="128 aa, similar to sp:S3AD BACSU STAGE III
SPORULATION PROTEIN AD from Bacillus subtilis (133 aa);
34% identity in 103 aa overlap. S.D. unclear. Putative
N-terminal signal sequence and 3 putative transmembrane
regions were found by PSORT.
CPE1830"
/codon_start=1

```

Query Match 7.8%; Score 144.8; DB 1; Length 301550;  
 Best Local Similarity 52.3%; Pred. No. 4.7e-22;  
 Matches 320; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

```

/transl_table=11
/product="stage III sporulation protein AD"
/protein_id="BAB81536.1"
/db_xref="GI:18145494"
/translation="MDVVIQVTSFVLISLFLVLKDNKSTPAVFLLLGAGIVILFV
MPVNEIIAPMODIARESMDSSYIAIVWKIATAIYLSFTFCSYICNDAGVTLGKKVE
FYKIMILLALPIMANILNLSILSIM"
gene
CDS
complement(3864..4061)
/gene="spolIIAC"
complement(3864..4061)
/gene="spolIIAC"
/notes="65 aa, similar to sp:S3AC BACSU STAGE III
SPORULATION PROTEIN AC from Bacillus subtilis (68 aa);
49.2% identity in 63 aa overlap. Putative N-terminal
signal sequence and 1 putative transmembrane region were
found by PSORT
CPE1831"
/codon_start=1
/transl_table=11
/product="stage III sporulation protein AC"
/protein_id="BAB81537.1"
/db_xref="GI:18145495"
/translation="MSDLSLIFQIAGVIGVILVDKVLDSQSGKKEYATLANIVGVII
LTMIIQLISRLPSSVKSMPFLF"
gene
CDS
complement(4137..4652)
/gene="spolIIAB"
complement(4137..4652)
/gene="spolIIAB"
/notes="171 aa, similar to gpu:AP001516 225 mutants block
sporulation after engulfment from Bacillus halodurans (170
aa); 34.5% identity in 145 aa overlap. Putative N-terminal
signal sequence and 1 putative transmembrane region were
found by PSORT
CPE1832"
/codon_start=1
/transl_table=11
/product="stage III sporulation protein AB"
/protein_id="BAB81538.1"
/db_xref="GI:18145496"
/translation="MKVFFLLIIVLSSLIYLYGFGFRNLSQLRELKRALIDFEND
IVYVTPLESIESIALKAKSPIKELFNEISFKLKNNEVENVYMAKESINEHKENN
LRNDFEILLDSKSGEINVEGQIKIFNLAKELDIELEIAEDECNKNTKVIYLGV
AVGAMTAFIV"
gene
CDS
complement(4799..5683)
/gene="spolIIAA"
complement(4799..5683)
/gene="spolIIAA"
/notes="294 aa, similar to sp:S3AA BACSU STAGE III
SPORULATION PROTEIN AA from Bacillus subtilis (307 aa);
40.3% identity in 293 aa overlap
CPE1833"
/codon_start=1
/transl_table=11
/product="stage III sporulation protein AA"
/protein_id="BAB81539.1"
/db_xref="GI:18145497"
/translation="MKEIFAVLPDKINSLCKDKSNLKLQEIIRVKVKPLNIVLDNTE
TIYVYIREDVKALIQKISNLSYAFEDIRQGYITIOGHRVGLAGOCVIEDNSIK
TIRNTSNIIRVCREIIVGGSNLSVENNRVNTLIISPPKCGKTLTLDRTIRNIS
NGSIOGLGKKTVDVIDERSEAACYNGIPQNMVGRDIDVYDNCIKSEGMMMAVRGLS
PEVITCEIDGTYKDMGLMAYNSGVSIATLHGRNVEELYRPRVPFREVENNINKV
VLSGKKGTGTEGIVNV"
gene
CDS
complement(5882..6244)
/gene="CPE1834"
complement(5882..6244)
/gene="CPE1834"
/notes="120 aa, no significant homology"
/codon_start=1

```

QY 134 AAAAATCATGCGAGTGTCTCAGGAAATTTTACCACCGTGAAGTACCAACTCGCCCA 193  
 Db 122306 AATGTGAAAGCTATATTTACAGAAATAGATCAATGAAAGAGCTGAAGAGTATCTGGA 122247  
 QY 194 GATTAAATGTGAGGAGCATGTCTTTGTCCCGAGGTTTCAGTGAAGAGCATCTCT 253  
 Db 122246 ATATTAATAGTGTGAGCATTTACAGATTCCAGTAAAGCATGTTCACTGTAAGT 122187  
 QY 254 TCTGATCTGGGAAAAAACAATGACACAGAGCATTTATCTGACATGCTGTGGCA 313  
 Db 122186 GCTCAATATGAGTCAAAATGCTCTTCAAAATGCTGAAAGCAGAGCAATAGTGTAGCT 122127  
 QY 314 ATGCTTATGTTTGTGATGAGGATATATATGATTTGAGGCGCATGCTGGGAGCT 373  
 Db 122126 ATAAATATCTTATTTATGATACATACTATAGAGTTCCAGATTTATTTGACATGATGCA 122067  
 QY 374 GTTCTTCTGAAATCTTTGCTATCTATCTGAGGACCTCTACATTTTGGATGCCCACTACC 433  
 Db 122066 CTAAGTCTTTATATATTTATGAGTCTTATATTTGATAGTATGAGTAACTCTTACT 122007  
 QY 434 TTGTGAGACTCGCTGGGATGTCTTCTGCTATGAGGATGCGCGTATGATGCAATGTCTT 493  
 Db 122006 TTACAGGATATAGCAGCTTTCTTATTAACAGTATGAGTATGAGTATGATGCTAACTACT 121947  
 QY 494 GTATTCGAAAGATCCGAGAGGAATTTTATTTGCTCAAGTCTTAATAAATCTGTAGAA 553  
 Db 121946 ATATTTGAAAGATCAAGAGAACTTGAATGCTAGAAATGATATTAACATCAGTATGAAA 121887  
 QY 554 AAAGATATACCAAGGCTTTTGGAGCCATTTTGTATCTTAACCTGACTACAGTATTTGGCC 613  
 Db 121886 ATAGGTTTATGATGCTTTTAAAGTCAATATGATTTCTTAATATACAACTTATTTGCT 121827  
 QY 614 TCAGCACTCTTTCTTCTCCAGATACAGGCGCTATTAAGGGTTTGTGATGATGATT 673  
 Db 121826 GATTAGTCTATACCTCTTGGATCAGAGCAGATTAAAGATTGCTTAAACTATATTA 121767  
 QY 674 TTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCAATGATCAATTTTCTCATGTG 733  
 Db 121766 ATAGGATATGAAATGATGATGTTTACCGCTATTAATTAATGACAAATCTTATGATTTA 121707  
 QY 734 TGGATGAATAG'745  
 Db 121706 GGATTCACATG 121695

RESULT 11  
 AE007729/c 12592 bp DNA linear BCT 27-JUL-2001  
 LOCUS Clostridium acetobutylicum ATCC824 section 217 of 356 of the  
 DEFINITION complete genome.  
 ACCESSION AE007729 AE001437  
 VERSION AE007729.1 GI:15025281  
 KEYWORDS  
 SOURCE Clostridium acetobutylicum.  
 ORGANISM Clostridium acetobutylicum.  
 Clostridia; Clostridiales; Clostridiaceae;  
 Clostridium.

REFERENCE  
 AUTHORS Nolling, J., Breton, G., Omelchenko, M. V., Markarova, K. S., Zeng, Q.,  
 Gibson, R., Lee, H. M., Dubois, J., Oiu, D., Hitti, J., Wolf, Y. I.,  
 Telusco, R. L., Sabatne, F., Doucette-Stamm, L., Soucaille, P.,  
 Daly, M. J., Bennett, G. N., Koonin, E. V. and Smith, D. R.  
 Genome sequence and comparative analysis of the solvent-producing  
 bacterium Clostridium acetobutylicum  
 J. Bacteriol. 183 (16), 4823-4838 (2001)

TITLE  
 JOURNAL 21359325  
 MEDLINE 11466286  
 PUBMED 2 (bases 1 to 12592)  
 REFERENCE Childress, D., Zeng, Q. and Smith, D. R.  
 AUTHORS Direct Submission.  
 TITLE Submitted (24-JUL-2001) GTC Sequencing Center Production,  
 JOURNAL Finishing, and Bioinformatics teams; Genome Therapeutics Corp., 100  
 Beaver Street, Weltham, MA 02453-8443, USA

FEATURES  
 source  
 gene  
 CDS  
 gene  
 CDS  
 gene  
 CDS  
 gene  
 CDS

Location/Qualifiers  
 1..12592  
 /organism="Clostridium acetobutylicum"  
 /strain="ATCC 824"  
 /db\_xref="ATCC:824"  
 /db\_xref="taxon:1488"  
 complement(83..985)  
 /gene="CAC2277"  
 /note="secP"  
 complement(83..985)  
 /gene="CAC2277"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Preprotein translocase subunit SecP"  
 /protein\_id="AAK80234.1"  
 /db\_xref="GI:15025282"  
 /translation="MLKYVHKHWFGISLTIIPLGIAFMIVRGVTTGKPLEFGIDPV  
 GNSMDIKIGSFNKTVDYKIVDKYIKDSTITVNNTEVEIKSMTDAIGNKIFDM  
 OKKYNKSKALVSOEIKGTLGSELKTXALIALVALVILYIAMRPFRRGVAMI  
 SLVHVLVTLAYVYGFALPLNPTPLAGMLTLYGSMDSPTIVFPDRIRSRGRHMRST  
 SEIADPSINETLRISITVTLVTLVAVSVHVFPSVRDFTKLVLGVISGCVSSIFIA  
 SPLWVIFKERRANKKLOIKAEAK"  
 complement(985..2238)  
 /gene="CAC2278"  
 /note="secD"  
 complement(985..2238)  
 /gene="CAC2278"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Preprotein translocase subunits SecD"  
 /protein\_id="AAK80235.1"  
 /db\_xref="GI:15025283"  
 /translation="MKKGRSTIFPFIISVLIAGLAVLGSTGLANGDRIKSPFSID  
 KQILDQGVSVVEEIQKVNSDAMRRTLEMLRVNKGVSSTVSRRGNKIEIDVP  
 GMYNDEVINKIGTKLFVDPDKTTLTSDVKKATAGTDPNTNPTIISQLNSG  
 TKKPADATOKPLFGOKISIIYMDPLSDPVDVITGNAQIIGNKSIOBAPQIANIN  
 SGALPDLTKVNSKTVGSLGASALPNSTIAGAVACIIELFMILVYRIPGMADIAL  
 VLFVVLGAFYPAVAVTLLTSLGAGLLITGANVANDVILFRFRBEELKGTVSAP  
 DAGFHKMSIIDSNTTITISGVILPALSGEVKGFALTVLVIGVLISMTALITVRL  
 LMAIIMGLINKASHFVNLESGSK"  
 complement(2493..3857)  
 /gene="CAC2279"  
 complement(2493..3857)  
 /gene="CAC2279"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Heme biosynthesis (hcrJ-2) family protein"  
 /protein\_id="AAK80236.1"  
 /db\_xref="GI:15025284"  
 /translation="MSLVKFTODDORYIIDVNSGSIHEVDENVYDLIDEDKIRSKK  
 LLDRLNKYKEEISAVDEIKELIDEGVYSKLYEDIAKSTSPSTIKALCLNTA  
 HDONLCKYCFADBEYKGCRIEMSEVAKSIDVYIKSGPRKNIENVLFGGEPLMA  
 FDIKQVVEYAKKEEKYKNIRFTWTNATLNDIEMDYLDKMGNIILISDGRLEV  
 NPKVRVADGTGSHDAIKNIRKVARSDSKOYVYRGFTLRNNDVFEADVADMG  
 FKEISIEPVLDPDNLISREDDIQIPEQYDKLYKMDTKRKEKEIKFPHNDIN  
 GGPCTYKRLSGGAGHEVYSITPSSGVYFCHQPVGNEDPKMTIYDDINENMGIER  
 RSHIYNKPKRCRCWAFYCSGGCCQANNYFNGDMHPIYEIGCKQKKEICAIALKA  
 QWVD"  
 complement(4317..4688)  
 /gene="CAC2280"  
 complement(4317..4688)  
 /gene="CAC2280"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Predicted membrane protein"  
 /protein\_id="AAK80237.1"  
 /db\_xref="GI:15025285"  
 /translation="MEKSKTIYVSAAGVLRALPFTLITLIFSPVSTKISSEGITN  
 WVLVTTLTSLVWYSGYSKSGKGLNGLVGLFYILIFIVISLVDSGSGMQLRD  
 FIRVTLAIVGLISGLGINI"  
 complement(4778..5062)  
 /gene="CAC2281"  
 complement(4778..5062)

```

/gene="CAC2281"
/codon_start=1
/transl_table=11
/product="Uncharacterized secreted protein, YajC family"
/protein_id="AAK80238.1"
/db_xref="GI:1502528"
/transl_table="MPNITVLILLYVFLNGFYVYIPIPRAKKKKXNMKLMSEMRVND
KVITGGIMGQITNIQIOFVLIITQDPDKVIKIKKEAIAVETRAKNQK"
/complement(5095..6225)
/gene="CAC2282"
/note="tgt"
/complement(5095..6225)
/gene="CAC2282"
/transl_table=11
/product="Queuine tRNA-ribosyltransferase, tgt"
/protein_id="AAK80239.1"
/db_xref="GI:1502528"
/transl_table="MYTILKKTGKVRGRFETVHGTTETPVFMVNGTLAAIKGAVSSM
DLKINQOVELSNTYHLSRPGDDVVKKLGLKPMNWRPILDTGGGQVFSIAEIR
KIKEGYVFNSHIDRKIFMGPEESMKIQSNLASTIAMAFDEVNPSPREYVERSVE
RTTLMVRCKEELRLNSLPOTINKQMLFGINQGVYDDIRIEHAKTIAKMDLDGVA
IGGLAVGETHEEMVRIIDAVVPHLPQDKPIVLGMVGPSPNILEAVSRGVDFDCVLP
RNGRHGVFTFKHKNLNEKPELDASPIDEGCQCTCKHYSRAYIRHLFKAKENLAM
RLCVHNLNFTKLMEDIRNAIDADCFEFPKNEKLKAWAGKA"
/complement(6243..7268)
/gene="CAC2283"
/note="queA"
/complement(6243..7268)
/gene="CAC2283"
/transl_table=11
/product="Queuine synthetase, queA"
/protein_id="AAK80240.1"
/db_xref="GI:1502528"
/transl_table="MQLEDFYIYPELIAOHPCKDECRMLVVDKKSGEVHRIFK
DVIDLPDGLVLDNRVLPARLIGNKVGTTGGKMEFLLRIDKIDKTWETLVKPKRA
KVGAFFGGVGLAAVTKSIGEDGNRVVEFYEGIFEEILDKLQMLPPYITBELEN
KEEYTVYKNGKSAAPTAGLHFTNELLDRIRAKIKTVPFLHVLGIGTPRPVKVEN
IESHHMSEYMLKESADTINETKNGGRIISVGTSTRLETIADEKGNVRQSGW
TDIYIPGYKYKIVDGLITNHLPKSTLVMLVSALSSEDNINLAKYKLAVERNRYFFSF
GDAMPIK"
/complement(7305..8354)
/gene="CAC2284"
/note="ruvB"
/complement(7305..8354)
/gene="CAC2284"
/transl_table=11
/product="Holliday junction specific DNA helicase, subunit
ruvB"
/protein_id="AAK80241.1"
/db_xref="GI:1502528"
/transl_table="MDRILTSVNLEDSABYNRPQKLNIEYIGSGKVKERMDIFIRA
AKKGEALDHLVFGPPGLGKTLIANIAKEMGNLKVTSGPATIERAGDLAAITGLS
EYDLFTDEIHLRSVEEILYPAMEDYALDIIIGKAAKSIIRLPLKFTLIGATFR
VGLLTAPLRDFRCPMEFYNEDELKEIIVRSSILDIDIDEADAIEIAMSRSRGTFR
IANLLKRVRYADVCKRGIYNLDSAKNALNLEVDDEGPDSDINKLIIQAIIVNNFAG
PVGIETLSYFVGEIDTIEDVYEPYLQKGFIMETPRGRMATKAYEHLKVPFNKKQ
GTVNDQCSFFKKEK"
/complement(8367..8972)
/gene="CAC2285"
/note="ruvA"
/complement(8367..8972)
/gene="CAC2285"
/transl_table=11
/product="Holliday junction specific DNA helicase, subunit
ruvA"
/protein_id="AAK80242.1"
/db_xref="GI:1502529"
/transl_table="MYEYIKGYIKINKDYIIVDNGVGYYKIFTSGMTMANMPKPEED
VLIIYLTQIVRODFGLYGLTKOBELENFKLLTINGGAKASLSLSICTVNNLKYAI

```

```

ITGDEKTIKAPGIGKKTQRIILLEKDKPKSTDISKGNSEINNLDVVDYDEHSKKLEE
VRPALNSLGYSEKETDRAINNVKDSGIEINIIKSLRFLMN"
/complement(9028..9675)
/gene="CAC2286"
/complement(9028..9675)
/gene="CAC2286"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein, similar to protein from
Clostridium histolyticum (GI:3892648)"
/protein_id="AAK80243.1"
/db_xref="GI:1502529"
/transl_table="MTLKKVILSVALLVTFIPSCIGYRGNLKIIMNAEKYKISIG
IDEFLGRNESEKARYTSNNKGGKTLALLIVKQNYDITFDPEDETKNNVLTSSBETNA
EKHELAGMTKEQITRSYEQYIYIWNQMKVYIKNRYINNHNKVDQGVFLNKDKNSD
KINVYKVDNNGSLIKTNEQAYISVNQLNQNETRDLRLKRGFLVFGSKEDAKKLAGYD"
/complement(9880..11235)
/gene="CAC2287"
/complement(9880..11235)
/gene="CAC2287"
/codon_start=1
/transl_table=11
/product="AcyL-CoA reductase LuxC"
/protein_id="AAK80244.1"
/db_xref="GI:1502529"
/transl_table="MKLEFTCGEEITIKVYNKLKLEMTSGKGDVEKELERLKNKNK
VHDIKTEETIDLLDRCAKLWDDNYSKKHITETLAQITNQSFELVNTVEFKSMOMLLRE
NIQNTIKRELGEVDILDSHWKTNVGYEHRQPRGIIFHNISGNAPVVPIMISMOLLK

```

Query Match 6.8%; Score 127; DB 1; Length 12592;

Best local Similarity 50.5%; Pred. No. 5.9e-18;

Matches 335; Conservative 0; Mismatches 325; Indels 3; Gaps 1;

Qy	198	TAAATCTGGAGCATGCTTTTCTCCGAGGTTCTCAGTGAAGACGATCTCTCTG	257
Db	1589	TAACTCAGGAGCTCTCTCTTAACACATAAAGTTGTTCAATCCAAACTGTAGTGCT	1530
Qy	258	ATCTGGGAAAAACAATGTACACAGGCATTCTCAGCATGCTGTGGCTTGGCAATGC	317
Db	1529	CACITGGTGAAGTGCATCTCCAAACAGTACTACTTCAGGAGCAGTTGCTATATGATAA	1470
Qy	318	TTATTGTTTGTATGAGCGTATATTATAGATTGGAGGGTCATCGCTCGGAGGCTGTT	377
Db	1469	TATTCTTATTTCATGATTTTAGTATATAGATACCTGCTTATGGCAGATATAGACTAG	1410
Qy	378	TTCTGAATCTTTTCTTATCTGGCAGCTCTACAGTATTTGGATGCGCCACTCACCTTG	437
Db	1409	TTTTATTACTGTTTGGTTTGTAGTGCAATTTAGCTGTAAAGTTTACATTAACACTTT	1350
Qy	438	CAGGACTCGCTGGGATTTCTTCTCTATGGGATGGCGGTAGATGCAAAATGTTCTTGAT	497
Db	1349	CAGTATAGCTGGACTTTTACTTACGATAGTATGGCTGTAGATGCCAATGTTCTTAT	1290
Qy	498	TCGAAGAATCCGAGAGGAATTTTATGTCTCAAAGTCTTAAAAAATCTCTAGAAAAAG	557
Db	1289	TTGAAGATTTTAAAGAAGAACTTAAAAATGGTAAAAACAGTTAGATCAGCTTTTGTAG	1230
Qy	558	GATATACCAAGCTTTTGGAGCCATTTTGGATCTTAATCTTAATCTACAGTATGGCCTC	617
Db	1229	GTTTTCATAGGGCTATGTCTCATCAATAATAGATTCAAATGTTTACCACCTATTATT	1170
Qy	618	CACCTTCTTTTCTCTAGATACAGGGCTATTAAAAAGGTTTGTCTTGACATTTGATTTAG	677
Db	1169	TTATACITTAGTCTTGGAGAGTGGCGAGGTAAAGGATTTGCGTTAACACTGTTAATG	1110
Qy	678	GAATTTTCTCTCAATGTTTACGGCTCTTTTTCATGACTAAATTTTTCATGCTGGA	737
Db	1109	GTGTTCTTATAAGTATGCTACTGCTATTATTGTAACCTAGAAGACTTTTGTATGGGCTA	1050
Qy	738	TGATAAGACCC---AACATACACAGTTGCATATGATGAATAAGTTCGTGGGGATAAGC	794
Db	1049	TAGCATGGGACTTTATAAATAAGCATCTCATTTCCGAGTTTAATCTCGAAGGGATGGTA	990

QY 795 ATGATTTCTTGAGAGATGCAAAAACCTTGGCGCTTTCGSAAGTTTCTTTAG 854  
 DB 989 AGTAATGCTTAAGGTTTATTAACATACTAAGTTGGTTGGTATCTTACTATAAT 930  
 QY 855 GTT 857  
 DB 929 ATT 927  
 RESULT 12  
 AP005369/c 300600 bp DNA linear BCT 17-AUG-2002  
 LOCUS Thermosynechococcus elongatus BP-1 DNA, complete genome, section 1/9.  
 DEFINITION Thermosynechococcus elongatus BP-1 (strain:BP-1) DNA.  
 ACCESSION AP005369 BA000039  
 VERSION AP005369.1 GI:22293721  
 KEYWORDS  
 SOURCE Thermosynechococcus elongatus BP-1 (strain:BP-1) DNA.  
 ORGANISM Bacteria; Cyanobacteria; Chroococcales; Thermosynechococcus.  
 REFERENCE  
 AUTHORS 1 Nakamura, Y., Kaneko, T., Sato, S., Ikeuchi, M., Katoh, H., Sasamoto, S.,  
 Matanabe, A., Iriyuchi, M., Kawashima, K., Kimura, T., Kishida, Y.,  
 Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Nakazaki, N.,  
 Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M. and Tabata, S.  
 TITLE Complete genome structure of the thermophilic cyanobacterium  
 Thermosynechococcus elongatus BP-1  
 JOURNAL DNA Res. (2002) In press  
 REFERENCE 2 (bases 1 to 306600)  
 AUTHORS Kaneko, T.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-JUN-2002) Takakazu Kaneko, Kazusa DNA Research  
 REFERENCE Institute, The First Laboratory for Plant Gene Research, 2-6-7  
 AUTHORS Kazusa-kametar, Kisarazu, Chiba 294-0812, Japan  
 JOURNAL E-mail: kaneko@kazusa.or.jp,  
 URL: http://www.kazusa.or.jp/cyano/Thermo/  
 FEATURES  
 source Location/Qualifiers  
 1. 300600  
 /organism="Thermosynechococcus elongatus BP-1"  
 /strain="BP-1"  
 /db\_xref="taxon:197221"  
 /note="BP-1"  
 129. 1142  
 /gene="t1r0001"  
 129. 1142  
 /gene="t1r0001"  
 /note="ORF ID: t1r0001"  
 probable oxidoreductase"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAC07554.1"  
 /db\_xref="GI:22293722"  
 /translation="MKAVATABEGEPDVLVLRQVPEPQIEADEVKIQLRASVNPID  
 TKLRQCTPPDRPALIGDGAAGVAVVAAARPRVGEVFCVGGIDRGCVAR  
 YAVPEAAVAHKEKTSFIQAAALPLAVITAMEALGDRGVPPPLNVISTKTVLIHNS  
 AGGVGHIAIQARRAQAQVATTISPPAKQFTALGTLAINVTIDWVAVLDWTS  
 KGVLDLADIVGATFAETPKAVRPGVGLATLLEPAVDTPMKIARORNLILQTLMLTP  
 QLGIREALAHQGIILEQVATLVDRGELRVVDKTFPLGAADAHRYLSQRLVGKVV  
 Lip"  
 complement(1351. 2331)  
 /gene="t110002"  
 complement(1351. 2331)  
 /gene="t110002"  
 /note="ORF ID: t110002"  
 /codon\_start=1  
 /transl\_table=11  
 /product="serine proteinase"  
 /protein\_id="BAC07555.1"  
 /db\_xref="GI:22293723"  
 /translation="MVABAGPAAVVSIDTLRLDRGSEDPPLAPFPVDPVPLRQGGSGF  
 IFTPDGKIMTAHYVEGASAVRYTLPPRGQVKGVLGADSLTDVAVQINAKLPTQ  
 LGNSDTLPGEMATAIGNPLGSLNTVTAGIISANGRASSELGAADKXVSF1QTDALIN

PGNSSGPLNAGGVVGNVAVISOAGLGAIPINTAVGLAEQIITTPRAOHLVGI  
 BMVPLPELALQIREQGNMTLNRQGLIIGVANSPPAAKAGLOAGMIKVNINO  
 PTPQGVGVGRTKRGKRTIEIRGDRQRLRLKRPMPBMYSAQPE"  
 complement(2402. 2578)  
 /gene="t110003"  
 complement(2402. 2578)  
 /gene="t110003"  
 /note="ORF ID: t110003"  
 unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAC07556.1"  
 /db\_xref="GI:22293724"  
 /translation="MAADVQRNHVQNKARVDLMSWVEYGPQAAFPWDGDRPWRPN  
 GELINPASHPLP"  
 2577. 2762  
 /gene="t1r0004"  
 2577. 2762  
 /gene="t1r0004"  
 /note="ORF ID: t1r0004"  
 unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAC07557.1"  
 /db\_xref="GI:22293725"  
 /translation="WLTIVISYIEQPMFTFDSFFGPVTLQPRNENVDERRMRCKTHNA  
 DLQALMKGLIYVEELG"  
 2816. 4534  
 /gene="urec"  
 2816. 4534  
 /gene="urec"  
 /note="ORF ID: t1r0005"  
 /codon\_start=1  
 /transl\_table=11  
 /product="urease alpha subunit"  
 /protein\_id="BAC07558.1"  
 /db\_xref="GI:22293726"  
 /translation="MSYRIDRQTVAEYGPVIGDRLRLADTLIIEIHDYTHGDEV  
 KRGGKVRIDGKQSPINABEAGAVVITNNVILPMKGVKADVGIKGKIYKIKAG  
 NPYTGEDVILIGPTEAIAGEMLITNGGIDAHITFCPOQIATAIAGITTMGGG  
 TGPATGNTATCTPGPWNLYRMLQADAFPVNLGFLGKNSQPGGLEQVAGVGL  
 KLHEDMGSTPNAIDTCLSAEDYDIOVAIHDTPLNESGVEEDTIAFNKRTIHAVTE  
 GAGGGHAPDIKLCGOANVLPSTNTPRTYVNTLDEHLDMLMVGHLDP3LEDAVAF  
 ABRIRIRRTIAEDTLDHGARSIISSDSOAMRGVGEIIRTTQTAHMKVORGLRD  
 PORPGVDHNPRAKRYAKYTINPAITHCIAIYVSVEGKADCLKRPAREGKPE  
 LVIKGVLAIVAQMDANASTPTPQVPHMQPMFASGGCTTTSVTFMSQAGIANNIPB  
 QLGLOKTVLPVGHIRQLRKADKLDVLPHEIVDETYEVADGELLTCEPATVPLA  
 ORYFLP"  
 4538. 5032  
 /gene="t1r0006"  
 4538. 5032  
 /gene="t1r0006"  
 /note="ORF ID: t1r0006"  
 hypothetical protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAC07559.1"  
 /db\_xref="GI:22293727"  
 /translation="MTDAVSALFQOQQAQFERGNTQOSITTLAQAVALEGNTTOSGE  
 ISLMLVTASAGDQEAVALSLCROLQRHDPDHTROESRLILALEPQIKRRPWCSE  
 IIDLSHLGDRAVTSYRHSRKRPSAFTPQGPFPATPLPNAFTWIALTGAVATLL  
 AAML"  
 complement(5430. 7640)  
 /gene="t110007"  
 complement(5430. 7640)  
 /gene="t110007"  
 /note="ORF ID: t110007"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cellulose synthase"  
 /protein\_id="BAC07560.1"  
 /db\_xref="GI:22293728"  
 /translation="MSHEPPRPASPPQWVAPMWMGFVFSILLFGVLLATTWTRPTLF

```
PIDLVNPDQLRPLDILLQMBEDELVSVMPLVLVGMVAIALNFIPSNVSRLLIRFIVL
LFGKYLWVGWGTVDINDAHLMSFAASFGYGLVLEFFTYLLFYQYPAWLTITAWRSQQ
ANAYQRYLVWGEVCPYDIIFPTYNPPYILRRTIVACQAIHYRNKSIYVLDGRSE
IADICAGLVNLTPTNPNRKNALHAKHTTGELIAVFADADFIPOFNFLRTYVGF
FQDDVSMVOTPOHFNFVHSONLGEFNMWGDMEYFFGFIOGRDFGNAIICCGTS
VWVRRALEAGVYTCVVEDFOTGFKMOIAGVRLVYLNELISMGESPFODYLEQ
RLRWLOGNQIYFCGDDLPWLSKSWFORSCHLSLLLNINPFICTFLGPISLMT
GISUTVAILSEYLFYALPTILLMIATFSWATBEGRYFSLMGEVEVETPFGMVQLIKI
LRNPFKGISIVTNKGTLSNRKRLNRYTWPLVAFIAVSGVFIRYGGYJHLHWPM
EYRSGLEVMWLTLYNAFTALIAVLSIDQPSRRQDRFPVCTVCSFQIGDDQYWG
TRDVSETGAALLTAGRFIAQODTVGTLTFLEODFSVTAAVRSRTEEBQCCLYLRP
LEVSDEAQRGLVQLLYGLTWMHKPKAPNGLDAFWPMLIRLFDERSLSLSYSH"
complement (7667. .8746)
/gene="ycf81"
complement (7667. .8746)
/gene="ycf81"
/note="ORF ID:tl10008
hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAC07561.1"
/db_xref="GI:22293729"
/translacion="MLETILLIIIVMGIAVGNSIDLLPETVLAQVANVRGLQWMA
AFGAIVGIALGLLQSLYHLERSIROLPELTLSRAVGLVGLLNLANLAPIELLP
IPKDFSIKPLIAVLTSILFAYSGTTLADSHGPALLRLINPNVASSILAEGLMKPAR
AKVLDTSICIDGRLEALNGLVLEGQIIVFQVFLQELQIADAGNEQKMRGRGLDV
LNRLQASIDRIIVHSADYPELTVDAKLVRLOEINGTLVTNDLNLNKVAFQKVDV
FVNVELAQALPIYLPDGTLELKLKEGEPQAGVGYLEDGTMVVVEGVHDIGDQLS
VVTLSALQTSAGRMIFARLQMPWA"
complement (8796. .9758)
/gene="tl10009"
complement (8796. .9758)
/gene="tl10009"
/note="ORF ID:tl10009
unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAC07562.1"
/db_xref="GI:22293730"
/translacion="MATAETAPIAEIALVHVHTDRLRLRIQELKTDVGRDLSLTKTL
RGIOVHNPLAASITIEYHREQITPLQLAAIOFMGEVOILGQNGRGLONLRAFDLP
EPREVGNKLTAMGFLMGVGLDILGVMGTAGGLLGMGGVAMGVQVTFVGGVIG
ARLGEATVQNIQOFALTEPPIAKALEIHSQDKIGGAAGELAGLAGOVVLGPV
GETLGRVNVNMGSQLQEDIGRLAEPSPGPPPSLHFFLEVMWVTSQTFAKETLLA
TLGIVARGILGPQAEAGIRAGSRMRHLDROQSNTAAKEKY"
9717. .10886
/gene="tlr0010"
9717. .10886
/gene="tlr0010"
/note="ORF ID:tlr0010"
/codon_start=1
/transl_table=1
/product="ABC transporter ATP-binding protein"
/protein_id="BAC07563.1"
/db_xref="GI:22293731"
/translacion="MHELGNNRFSFGHNPQSQALVAVVARTKEGEISVQMLRSH
TELTMTATVTKVHLRKSYGAITAVADVSLTAHQEIGFVLGPNKSGKTLTLCITL
SRPDAGLEVCGFVLPQVHLPVROKGVVAQEVADKLITGQFLEQLAALYHPRAL
IPERIEAVLDLQWRDROCGTYSGIRKPLDLAGLHQPOVLVLDPTVGLDIE
SRQVINDVLDLKAQGLITLTSHYLEVDLLSRLVILDKGRVIASGSPPEELKANIG
GRVILVRREFTPQEAQRQALLSQTCKVKSALINANQGNLSNLVDTVDVIAAIIK
RALFEAGLPLFLAQSRPSLDVVDVYLAATGQTLDDLAALAAQSDSKOLKEVMQR"
10886. .11296
/gene="tlr0011"
10886. .11296
```

```
Query Match 6.7%; Score 124; DB 1; Length 300600;
Best Local Similarity 51.2%; Pred. No. 1.9e-17;
Matches 289; Conservative 0; Mismatches 275; Indels 0; Gaps 0;

Qy 142 TGCCAGTCTCTCAGGAATTTACCCACGTCAGTGAGCAACTCGCCTCAGATTAA 201
Db 183759 TGCCGAATCTCCGCTGCTTCACTGCCCAAACCGCCCAATGATCTGCCATTGCA 183700
```

```
Qy 202 ATCTGGAGCGATGCTTTTGTTCCTCCGAGTTCTCAGTGAAGAGACGATCTCTTCGATCT 261
Db 183699 GGGGGAGCGCTGCGGTTCCCTTAGAGTGGTGGAAAAACCGACGCTGGTGTCTCCCT 183640
Qy 262 TGGAAAAACAATGTACACAAGGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTAT 321
Db 183639 GGGGCAAGACAGTATTCGCCATAGCCTCTATGCGGGCTTGGCCGGCTTGGGTGGTGTCT 183580
Qy 322 TGTTTTCATGAGCGCTATATATAGATTGGAGGCGTCATCGCTTCGGGAGCGTGTTCCTCT 381
Db 183579 AGTCTTCATGTTGGCTTACTATCGCTCCCGGACTCATTTGCCGATATTGCCCTGCTTCT 183520
Qy 382 GAATCTTTTGTCTTATCTGGGAGCTCTACAGTATTTGGATGGGCCACTCACTTTGTGAGG 441
Db 183519 ATACGCCATTTTACCTATGCGGCTTCTCTGCTTTGGGGTGACGCTCACCTTCGCGGG 183460
Qy 442 ACTCGCTGGGATTTGCTTCTGCTATGGGATGGCGGTAGATGCAAAATCTTCTTGTATTCGA 501
Db 183459 GATTGCGGGTTTTATTCTCAGTATTGGCATGGCGGTGTGATGCCAACGTCCTGATTTTGA 183400
Qy 502 AAGAATCCGAGAGGAATTTTATTGTCTCAAAGTCTTTAAAAAATCTCTAGAAAAAGGATA 561
Db 183399 GCGCACTCGGAGAGGATTACGGGACGGAACCCCTCTACGTTCCGTTGNAAGCGGAT 183340
Qy 562 TACCAAGCTTTTGGAGCCATTTTGTATCTTAACCTGACATACAGTATTGGCCTCAGCACT 621
Db 183339 CGATCGGCTTTTGCAGTATTCTGGATAGCAACGTCACCACTTGATTTGCTGTGCTGC 183280
Qy 622 TCCTTTCTTCTAGATACAGGCGCTATTAAGGTTTGTCTTTGACATGATTTTAGGAAT 681
Db 183279 CCTCTTTGCTTGGGACAGGGTTTGTCCGCGGTTTGGCGTTACCCCTGCCATTGGTAT 183220
Qy 682 TTTCTCTCAATGTTTACGCGCTCT 705
Db 183219 TGCATTAGTATGTTACGCGCTCT 183196
```

```
RESULT 13
AY078352 1557 bp DNA linear BCT 04-MAR-2002
LOCUS Rickettsia typhi export membrane protein SecD gene, complete cds.
DEFINITION Rickettsia typhi export membrane protein SecD gene, complete cds.
ACCESSION AY078352
VERSION AY078352.1 GI:19110413
KEYWORDS Rickettsia typhi.
SOURCE Rickettsia typhi.
ORGANISM Rickettsia typhi.
REFERENCE 1 (bases 1 to 1557)
AUTHORS On'gele,E.A., Radulovic,S. and Azad,A.F.
TITLE Characterization of the Sec operon of the typhus group Rickettsia
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1557)
AUTHORS On'gele,E.A., Radulovic,S. and Azad,A.F.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-2002) Microbiology and Immunology, University of
Maryland School of Medicine, 655 West Baltimore Street, Baltimore,
MD 21201, USA
FEATURES
source 1. 1557
location/Qualifiers /organism="Rickettsia typhi"
product="export membrane protein SecD"
protein_id="AAL82790.1"
db_xref="taxon:785"
1. 1557
/codon_start=1
/transl_table=1
/product="export membrane protein SecD"
protein_id="AAL82790.1"
db_xref="GI:19110414"
/translacion="MQLPKWIKVPSIICVTATICALPNFIOINKFLPHDSVNLGL
DLRGGANLLDVDFDYLNDSDMENIADTKNFKREKIGYKVLVHRNNIQLVSRSTE
VLSKKLIHKIDPEIIIEVKNKIKLYSESRNLNLLNKVQESIEIVMRIDSTGT
KEPILQOQGNKHTLLQVPGVDNPSYLNILGKTAKLTFHLVDENANIEEAIKGHIPLD
```







```
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC65400.1"
/db_xref="GI:3322696"
/translation="MGIRSEGLFTEVQAEKRTYFNVKENRVGDFVQVQVSKNVDGA
GFDRHAVVPEEEMQKFLQGNFSLDFLEKNKRLHLRQARSHTGERKTVRK"
complement(3573..5471)
/gene="TP0413"
CDS
/gene="TP0413"
complement(3573..5471)
/notes="similar to GB:AE000783 percent identity: 35.60;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphoglucomutase"
/protein_id="AAC65401.1"
/db_xref="GI:3322697"
/translation="MSLLNSLITSLTPQAPQLHLSTBSPGMRKVFYAPTPERT
VLTQKADAVLSAQVALSYARFLKKVPKRVVIARSRPTGEIISICTAILHLEGC
SLRCVGVPETIARYENADAFIYISASHNPFIYNGFKFLGNDGVLSPTOAQOLSE
DFVAQCLQPASQKIREKHATLNASILEPLLYNKKHQSLSYDLFIREVANSKPE
TEHRAFFHRTQRAAYSEQGTPLSILIDFNGSARAASIDRRLESGLVALFSIAETP
GDIHRIVPSSSLTCAQALTEAATRGSSPEERSIAFGFVPCDGRGNIVYDOTL
NRVIPHQAVFALSVSSEICELSRNLRMQPHRAPPIALVNGPSTLSRVEAIAQLDV
HYFTEVGEAHLIEKAHLRLKRGVCVRILGEGSNGGSLHPAAVRDPLHTVFALLKL
LLRGDKQPLGFAWCAHSHQTERAHAFTISDLFASLPAYTTPTHDTALLQIRTT
DHALLKAYQRVFERRWAQEKDTLKRFGICSRPILYRGREQDAGDFSRADGGL
KIQYNHGHOAVAFVWVRGSKTEPIFRISADLKGNDGEAGEALLQLHADVMREARCA
LSCADTT"
5574..6929
/gene="TP0414"
CDS
5574..6929
/gene="TP0414"
/notes="similar to GB:L42023 SP:P44917 PID:1005967
PID:1220986 PID:1205130 percent identity: 43.24;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="D-alanine glycine permease (dagA)"
/protein_id="AAC65402.1"
/db_xref="GI:3322698"
/translation="MGVVEVWGMHVAHVSPNVVLLGTGCVTVCMKFPFVVRLLM
YVLKQITGGRRKGGGSEVSAFRAVATAALGSGNIVGVATAIAGGPGAFIWM
VTGIFGNTGFQSVLVAVYRRQDGRFVGKPMYLDGKGVPGAGVLASLFCFSV
IASFGIGMTQFNSVALVDFEDVFCVDRVTGAVMLVLGLVSGVGLKSIWVTVGMVP
GMALYYCVGVVLVCCNTRQYLVSGAPAGTAAGVGEFAGSVVRQIAIAGISRG
VAVNEAGLGTAPATAHRAITDHPVRQGLGFIFFVFTVMVSVTAFAILITGWQGG
ASGALIVQSFERGLPGICQYVGVSVPLFAFTMLGWLIVGERCAFLCGGRVCPV
YRAWLFPVTVGVAGVSLSSVWNISDAFNGLMALPNLVGLLFLARQGMRLREFFPAQOE
G"
6910..7443
/gene="TP0415"
CDS
6910..7443
/gene="TP0415"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="T. pallidum predicted coding region TP0415"
/protein_id="AAC65407.1"
/db_xref="GI:3322703"
/translation="MRSRKGRRMRMWSHENGVAHGNVGVGAKFLESRCVLEKAE
GGEICWGVIVKEDIRVKGEOHGFCLCAGMGDFVGTGEIRAGDRTEEAHVTKG
EGKVFRRQGVGGTTLRVLEKCRIVGMGMVGGPGVVRVFCARSAHSALSAKRAQAY
LAWMPSIGCISPRVCPA"
complement(7390..8748)
/gene="TP0416"
CDS
complement(7390..8748)
/gene="TP0416"
/notes="similar to GB:AE000783 percent identity: 45.01;
identified by sequence similarity; putative"
/codon_start=1
```

```
/transl_table=11
/product="signal recognition particle protein (ffh)"
/protein_id="AAC65403.1"
/db_xref="GI:3322699"
/translation="MFEQLSASFKRIVGALGGRTATISERNIQEAVESIKRALLDADVH
VVRVFNQVTOHQAGQTVLASVPAQOFIKIVHERLTAFLGEHRSLSHLKGPDPYOS
IILLGLQSGSKTTSARKLAAYLKDAGRSPLAACDHVRAAASQAQLAVLGHVIGPVY
QHALPHEQPPCALDTQAGALYARSHGNDVLIIDTAGRUHVDAALWQELILLKLTLPV
VETLVADALTGQTVVRIABEFAAVGISGVLKFSDFSDTRGGAALSLSKITQOPLLF
VGTGERPODFEPFHPERAAGRILMGDIIVSLVEKAQKAFDAQSHAKAKKTKQSHERT
LSMDLHLOTIEKMWPLSLVEMIPLGLAVASADALDARAFKROKAIQSMYQERDN
FLIIGSRRRRIAGSGTSVADVNRILKNFERMRTLMRKTAWQSRRAHPKGDTPYGWP
HR"
complement(8741..10420)
/gene="TP0417"
CDS
complement(8741..10420)
/gene="TP0417"
/notes="similar to GP:1651274 percent identity: 29.26;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="apolipoprotein N-acyltransferase (cutE)"
/protein_id="AAC65404.1"
/db_xref="GI:3322700"
/translation="MCTAPLVSSAASAVLLAFAPINPEFWLAGSVLGLALVPLVYG
FLSPAKKIVACSYGLFVALVHACSSFNLKNGFALFTLGASTVGFYFFVLPFGVAF
ACTLRQAQARACAFALVMTLWVVKSTGLIAYPWGTVPMTAHSLSHLIQIADITGVW
GSLFLPLANACVAESLHFFIKKRDVSPVPRWLMLTGCLYCLSLYAGVATLGAAPR
TLLALAIQONADPMDTXXSEKNTLTAIHLTETALOTAPPLTPPYRKEKTLTHAS
ARAPDMVWSESSLRYPYEQYRHVYNALPAARPSAFILTLGALLVGTPLRLSGNS
TKGGYANAVALLURPDGVAQVQKMDVFAEFIPWGHMTSVORLAQMLAGFSSEWTP
GPGPLRPHVPCAGSVRFATPICYEDAPPSCAALHTQGSSELLINLINDNSKTSASA
EMOHVYVSLFRATIELRTLLVSTNSGYTVVIGPEKTRAAFPLOATSAVLHVVPVVPV
VRTYARMRDWIVLICALIFFAEGVRMAVHTRHSTTQAESSLQIQIRGEHV"
10447..11640
/gene="TP0418"
CDS
10447..11640
/gene="TP0418"
/notes="similar to GB:U26401 SP:P51570 PID:1002507
PID:1929895 percent identity: 30.73; identified by
sequence similarity; putative"
/codon_start=1
/product="galactokinase"
/protein_id="AAC65405.1"
Query Match 6.5%; Score 122; DB 1; Length 12504;
Best Local Similarity 56.1%; Pred. No. 7.6e-17;
Matches 230; Conservative 0; Mismatches 180; Indels 0; Gaps 0;
Qy 302 TGTGGCTTGGCAATGCTTATTGTTGATGAGCGTATATTATAGATTGGAGCGGTCAATC 361
Db 1423 TGGGGGCTGTGCGCGGTGCTCTTATTATCTGCTGTTGATCAAGAAGCGGGCTAAAC 1482
Qy 362 GCTTCGGGAGCTGTTCTTCTGAATCTTTTCTTATCTTGGCAGCTCTACAGTATTGGAT 421
Db 1483 GCTTGTGTCGCGCAGCTTTTGAATCTGTATATCATGTTCGGTGTGCTTTCAGCGTTAA 1542
Qy 422 GCGCCACTCACTTGTTCAGACTCGCTGGGATTTCTTCTGATGGGGATGGCGGTAGAT 481
Db 1543 TTGACGCTTACGCTTTTCGAGTATCGCGGGATGATTTCTCACTATTGTTGGCAGTGGAT 1602
Qy 482 GCAATGCTTCTTGTATTGCGAAAGAAATCCGAGAGAAATTTTTTATTGTCTCAAGTCTTAAA 541
Db 1603 GCGNATGCTGTTCTTTTGAGCGTATACGGGAAGAACTTGGCTGGGCAAAAGCCGGGG 1662
Qy 542 AAATCTCTAGAAAAAGGATATACCAAGGCTTTTGGAGCCCAATTTTGTATCTTAACCTGACT 601
Db 1663 GCTGCTGTGTAGCGGCTTTGAGCGTGTGCTTTTGGGCAATTTATGGATTCAACAGTGA 1722
Qy 602 ACAGTATTGGCCTCAGACATCTTTTCTTCTAGATACAGGGCCTATTAAAGGTTTGCT 661
Db 1723 ACGTTTATAGACGCGCTTTTCTCTTTCGGTGTCTCGGTACCGGTCTCTTAAGGGTTTCGCA 1782
```

QY 662 TTGCATTGATTTAGAAATTTCTCTTCAATGTTACGCTCTTTGCAT 711  
 Db 1783 TACAGTTGGCTATCGGGGTGGTGTCTCCGATTTACGGCATGTTTCGT 1832

RESULT 15  
 RPX03  
 LOCUS  
 DEFINITION Rickettsia prowazekii strain Madrid E, complete genome; segment 3/4.  
 ACCESSION AJ235272  
 VERSION AJ235272.1 GI:3861033  
 KEYWORDS complete genome.  
 SOURCE Rickettsia prowazekii.  
 ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; typhus group.

REFERENCE  
 AUTHORS 1 (bases 1 to 279110)  
 Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sichenitz-Ponten, C., Alsmark, U.C., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H. and Kurland, C.G.  
 The genome sequence of Rickettsia prowazekii and the origin of mitochondria  
 Nature 396 (6707) 133-140 (1998)

TITLE  
 JOURNAL Nature 396 (6707) 133-140 (1998)  
 MEDLINE 99039499  
 PUBMED 9823893  
 REFERENCE 2 (bases 1 to 279110)  
 AUTHORS Andersson, S.G.E.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-NOV-1998) S.G.E. Andersson, Siv.Andersson@molbio.uu.se, Dept. of Molecular Biology, University of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN

FEATURES  
 source  
 1. 279110  
 /organism="Rickettsia prowazekii"  
 /strain="Madrid E"  
 /db\_xref="taxon:782"  
 63. 1045  
 /gene="RP479"  
 63. 1045  
 /gene="RP479"  
 /codon\_start=1  
 /transl\_table=1  
 /product="OCTAPRENYL-DIPHOSPHATE SYNTHASE (ispB)"  
 /protein\_id="CAI4934.1"  
 /db\_xref="GI:3861034"  
 /translation="MNTIVKIQNLKDEVOLNDLIISCLSDAELEIKYGVYLVAG GKRIRLLITITAKMPYKGNHNIKLSAVEPIHAALTDVVDNSTLRRPPTANV IWSKTSILVGDPLFSQSEFLKMAVSGCIKAMNTIARASVITSSEVQVLYKNERRII TIDBYQOIVSKTALFGLGACVGAIIAEVDVRSKVONFGRLLGTIFQVIDDLIDY LGSBKQVGNIGDPLEGKVTLPILFYHLEDDQKQMLNMLKSDKRTDDPVKIRD LMKHAIYNSTVYLSLENEANNLINKIPQNIKYKYLFSIRFIYRSY"

gene  
 1065. 1138  
 /gene="tRNA Arg (CCG)"  
 /gene="tRNA Arg (CCG)"  
 1065. 1138  
 /gene="tRNA Arg (CCG)"  
 /complement(1194. 1265)  
 /gene="tRNA Gln (TTG)"  
 /complement(1194. 1265)  
 /gene="tRNA Gln (TTG)"  
 /complement(1316. 3091)  
 /gene="RP482"  
 /complement(1316. 3091)  
 /gene="RP482"  
 /codon\_start=1  
 /transl\_table=1  
 /product="unknown"  
 /protein\_id="CAI4935.1"  
 /db\_xref="GI:3861035"  
 /translation="MRDRINLKLKFLFENYNGYIIPSDKNYNEVVPYAKLEYI TGTGSSGIAIICDAAFPPTDGRVYEOANKEDLAFYKTYDLKEIKTLDKNIKIGY DPQLFTYQVLANINIFHKINENIVDKIWNKPILEPNSKYIYLHDIKAGVSHNDIKRK

CBETILSSSVTVGCKNNDDILVILDSASICMLNLRLASDVNTPLMEKAVILITSTK  
 IYLFINPRIDPEILINABPEITILPEKEENILPESKRKYLIDDSITSVHIMDLANK  
 KYKATVEPCLLAKACKNDIEIHAIDFHKDAVALCEPFAEFLYHSSNNVCSHSH  
 ELITESHSLCKLITAGAKOEGYVSDSFHICGFQNSAI IHYRANPKAKKIEGHIL  
 LIDSAQYAGATTDYTRTIIIVGIPCEQKRYOVYKSHIALTRAKFPKNIVTGANLD  
 ILARQYLMQDMIDYDPTGHGWSFVSGEPOSINLKNHILKAGMILSNPGYIIP  
 GKYGRIENLIYKENNGMLPEFETSLVPYAKLIDVALNLINDEINTIKEYNKRIRAK  
 IYNILSTQAKWNLINFLQSL"

gene  
 complement(3174. 4463)  
 /gene="RP483"  
 /complement(3174. 4463)  
 /gene="RP483"  
 /codon\_start=1  
 /transl\_table=1  
 /product="PUTRESCINE-ORNITHINE ANTIporter (pocE)"  
 /protein\_id="CAI4936.1"  
 /db\_xref="GI:3861036"  
 /translation="MSOKLSFWAVPALVTGSGITGVFLPLSLAPFGVYSIMWILS LFGMSIALVFSCICTKPEKPGPHVYVRSAPSGTILAPFTGWTWIIISVTSIVIS AICYLTPPKSOTIILDIQLILAAIATLNKGGKIKGKVEFTLTKFVPLVGL ALVFNHNDIVITAKVENFTPTMGVALLTFMGFTIGIBCATTTAGVADPAKTIIP RALIIIGCCVAVLYINSIGIIGLIPABELINSKAPVADAAFLRGGTWSKVITYAS VICIIGLNAWVLTSGOIALGLAEDGLPFPFAKNSNNAPTYGIIISCIQITPLLEFT SNNPAKQITQIIDPSVIAFLFVYLICSLAFLKVISSKENSYYLFPVALISIFCT WVIYKTPPETLIACSFITLIGIPVYGYWRLMNRKX"

gene  
 complement(4633. 4965)  
 /gene="RP484"  
 /complement(4633. 4965)  
 /gene="RP484"  
 /codon\_start=1  
 /transl\_table=1  
 /product="HESB PROTEIN (hesB)"  
 /protein\_id="CAI4937.1"  
 /db\_xref="GI:3861037"  
 /translation="MKNVISTLDAAKQVYKLIBRKAFTPGIRVINGSGCAGQTVY VYADNKNQFDEVEKEGVRIILIDPKLMIYILGSEMDVYETNFKSQFPTNERNANC GGKSGPSV"

gene  
 complement(4969. 5364)  
 /gene="RP485"  
 /complement(4969. 5364)  
 /gene="RP485"  
 /codon\_start=1  
 /transl\_table=1  
 /product="NIFU PROTEIN (nifU)"  
 /protein\_id="CAI4938.1"  
 /db\_xref="GI:3861038"  
 /translation="MAYSKVVIDHYENPNVNSLDKKKNVGTGLVGA PACGDVAKLQ IEVGDEIITDAKFPFGCGSALASSLVTEWI KGSYEDAKEIKNTEIAKELSLPV KIHGSLIADAIKAAIADYKORENKKS"

gene  
 complement(5430. 6662)  
 /gene="RP486"  
 /complement(5430. 6662)  
 /gene="RP486"  
 /codon\_start=1  
 /transl\_table=1  
 /product="NIFS PROTEIN HOMOLOG (sp11)"  
 /protein\_id="CAI4939.1"  
 /db\_xref="GI:3861039"  
 /translation="MNOQLKNTLPIYMDYSTPDIIDPVMEAMLPYFTTKGNPNSR SHSPGMEANAVENARSMVAKYIGADSKIIPTSGATESNNIVIGIAKFGYNNKKHII TIVSBEKCVLNACRHLBOGKITIYPLPKSGIIDLITLKAIVIGDITLAVVMVNN EIGVIOPLKEIKICRENRVFPNSDIAGFGKIPINNAECNIDLASISHKIYGRKGI GALVIRKRPVRVTPILINGGQGERNSGTLPTLPILVIGIASEIAYNMEKOTQIVN YLFDRLNNIHSGIIEVYINGDKQGRKGNLMSFAGVEGSEIIAIAKDLAVSSGSA TSASLEPSYVLSIGISEIELATSIRFGIGRPTTQELIDYAVNLVCSKIDKRLRSLPL MEMMOGVLDKKIRWTAH"

gene  
 complement(6740. 7870)  
 /gene="RP487"  
 /complement(6740. 7870)  
 /gene="RP487"  
 /codon\_start=1  
 /transl\_table=1  
 /product="NIFS PROTEIN HOMOLOG (sp11)"



AUTHORS Aoki, K., Oguchi, A., Hosoyama, A., Nagai, Y., Kuroda, M., Hiramatsu, K., and Kikuchi, H.  
TITLE Direct Submission  
JOURNAL Submitted (30-JAN-2001) Akio Oguchi, National Institute of Technology and Evaluation, Biotechnology Center; 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
COMMENT (E-mail: oguchi@nitech.go.jp, URL: http://www.bio.nitech.go.jp/, Tel: 81-3-3481-8423; Fax: 81-3-3481-8424)  
FEATURES On Jun 12, 2001 this sequence version replaced gi:13701258.  
source Location/Qualifiers  
1..301550  
/organism="Staphylococcus aureus subsp. aureus N315"  
/strain="N315"  
/sub\_species="aureus N315"  
/db\_xref="taxon:158879"  
/complement (249..947)  
/gene="SA1293"  
/complement (249..947)  
/gene="SA1293"  
/note="ORFID:SA1293"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="BAB42553.1"  
/db\_xref="GI:13701259"  
/translation="MSSTFTMLIYFVILMTLPLMAQHKVKSNTKYSQVNSTSGKTR  
HVALEILHANGIDVEVAKGEGFLTDHDKKVSLSPPANYPSPVAGTALIAEYV  
HA1ODHGGWFLFRALVPAVNSGLSTYIMLGIIITLPAISAFGSTLMTWAGAGM  
SLAVLFSITVLPVEFDASSRAMKOITALTINVEKEYHAKVLSAAMTYVATAVAV  
AELVRIITLVARSSD"  
/complement (1001..1588)  
/gene="SA1294"  
/complement (1001..1588)  
/gene="SA1294"  
/note="ORFID:SA1294"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="BAB42554.1"  
/db\_xref="GI:13701260"  
/translation="MTINTFMQYTLVQSRWLMMLICNLTGMTYGYIWMGEOLSTPMP  
HPKIFVDPSPALILYLVISITLILKONSILBALPVLTKYGMVAVNINLFIIEQ  
GDITVNGIYLMFSGISMAVQAIYVYPRKSMGISVAMTWFLNDIDVPHLOPPIY  
DFTTHVMQJGVSSCCSLVFGLLIYELNKLCKK"  
/complement (1578..2153)  
/gene="SA1295"  
/complement (1578..2153)  
/gene="SA1295"  
/note="ORFID:SA1295"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="BAB42555.1"  
/db\_xref="GI:13701261"  
/translation="MSEETLNQIKESFIEYLLFOYRPSRTAVWVNTIKYNEKLANI  
HFDTKINHELEIAEVSQSHASAIQTKKNIKMTNNEIFDTIANHNCAPFDIOIHPA  
NSKRREQRDLIVAQTESPSYQTYLHNSMAIDRHKALLIDVLLHNIDLSQNN  
EKORFYQIQLINTLKLWNKNOFEDLADD"  
/complement (2167..3411)  
/gene="SA1296"  
/complement (2167..3411)  
/gene="SA1296"  
/note="ORFID:SA1296"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="BAB42556.1"  
/db\_xref="GI:13701262"  
/translation="MEDIKYKLIIDINQIKLENDLSRVNEAITDNDALFILGETLVN  
GLMPQGLIEVRYLYHKYDPSEELLIFIGLMSENOYDALREYLSVSPSPKMLME  
ADLYQOINMMEVAIDKLOALELEBNPDIHPALAEYLYDGOYLRSYSEYVLENG  
EYONGNVLFSRMDCSLGQNSYDAIRLADENEDMETSVDYLKKAISYDKNDIQR  
AIKIMTTLISKPDYIQGYLYLQSLYENKKNYPALITGKEGLRLSQFYKELMTTTC

LEIEHGDANEGLLLKQALEVNDVAOEPLLISDLVYRNEDEVAETIIEILTYDEEDLD  
PFFPMWHLAPAYQOERDKQACHQEFELATPMTKNTIDPNSDIYFLIEIGQKQATLIL  
NOLILBPSNEMHDELSRLQS"  
/complement (3418..4716)  
/gene="aroA"  
/complement (3418..4716)  
/gene="aroA"  
/note="ORFID:SA1297"  
/codon\_start=1  
/transl\_table=1  
/product="3-phosphoshikimate 1-carboxyvinyltransferase"  
/protein\_id="BAB42557.1"  
/db\_xref="GI:13701263"  
/translation="MVSQQLIDISGPLKGEIEVPGDKSMTHAIMLASAETSNIYK  
PLUGEDCRINDIFRLGVDILKEDDKLVNSPGYKARFTIPQVLITGSGTIRLLA  
GLLSGIGIESVLSDVSIGKRPMDVLRPLKMDKNIIGIEBNYPLIIPKSVIRIN  
YQMEVASQOVKSAIILFASLFSNDITVILKELDVSRNHTEFMFHNPIPLEAELSTTT  
PDALDHKADPHVPGDISSAAFVIAALITPESVTHNVGINPTSGSIDIVKMG  
GNIOIPNOTTGAEPPTASIRIOYTPMLQPTITGELVPKAIIDELPVIALICQAVGST  
TKDAEBLKVKETNRIDTTRDMNLNLGFELOPNDGLIHPSEFKTNATVDSLTDRIG  
MMLAVASLSSSEPVKIKOPDAVNSPQGLPLKLLNEEG"  
/complement (4726..5790)  
/gene="aroB"  
/complement (4726..5790)  
/note="ORFID:SA1298"  
/codon\_start=1  
/transl\_table=1  
/product="3-dehydroquinate synthase"  
/protein\_id="BAB42558.1"  
/db\_xref="GI:13701264"  
/translation="MKLOTTYPSNNYPIFVHGALDHTSYIDQDQSLIDENYQ  
YFADRFDDLSTENYKVTIPAGEKTKTEBOYQETLEYILSHVTRKNTLIVGSGAT  
GDPAGFVAATLLRGVFIQVPTTILAHSSVGKGVKINSKQKNLIGAEPRPVIYD  
LDPLKTLPEQQLISGVAEYVKHALNGESTTOEIQHFEDRILLOSNGMDKYAKGI  
ETKLDIVADEKEQGVKRFNLNGHTFGHAEVYENHKIAGHAMIGIYIOFIVANILEN  
SNHDIQHYNYLTKIGYPLETITDIDFTIYVMSDKKNDQSGVMWLKHFSGDIV  
QHIDOTTLOHACEQLKTYRK"  
/complement (5816..6982)  
/gene="aroC"  
/complement (5816..6982)  
/gene="aroC"  
/note="ORFID:SA1299"  
/codon\_start=1  
/transl\_table=1  
/product="choistamate synthase"  
/protein\_id="BAB42559.1"  
/db\_xref="GI:13701265"  
/translation="MRYLTSGESHPOLTVIVEGIPANLEIKVEDINKMPKROGGY  
RRRMQIEKDTVEIYGVNNGYTLGSPITMYVNTDPTHRKMGAAPISEERENMK  
RTITKPRGHDLVGQMKYNNRDLNVLERSARETARAVGALCKVLLQDLIDYI  
SSVVEIGIKDQFDYDSETFKANLDNRVIDDSIAQAMRKIDAEKNGSGISGV  
QVVENMPVGSVYHYDRKLDGKIAQGVSNARKGVSFGGFAAEKPGSIODEI  
LYNSEIGYRGVSNHGLGEGSGNMPILVNGVMKPIPLTYPLNSVDINTKDEPKAT  
IRSDSCAVPASIVCEHVAREIKALLIEBQSHIEQLQQQIADRQLNVEF"  
/complement (7368..7568)  
/gene="SA1300"  
/complement (7368..7568)  
/gene="SA1300"  
/note="ORFID:SA1300"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAB42560.1"  
/db\_xref="GI:13701266"  
/translation="MKNISITILGILMFLKMLMELLYGAIFLDPKLPNPITKILFILTLI  
YIFVLVYKELIIFLKSXYNKA"  
/complement (7597..7722)  
/gene="SAS045"  
/complement (7597..7722)  
/gene="SAS045"  
/note="ORFID:SAS045"  
/codon\_start=1

```

/translation="hypotheical protein"
/protein_id="BAB42561.1"
/db_xref="GI:13701267"
/translation="MWLVIMIIHNNKARFLCCSKLEKFYSFLVTRIKRGIPFIK"
complement(7777. .8226)
/genes="ndk"
complement(7777. .8226)
/genes="ndk"
/notes="ORFID:SA1301"
/codon_start=1
/translation="hypotheical protein"
/protein_id="BAB42562.1"
/db_xref="GI:13701268"
/translation="MERTFLIMIPDAVORNLIIGEVSIRIERKGLKLVGGKLMQVPMEL
AETHYGHQKPNLDISITSPAPFAMVVEGEDAVNVSRIHIGSTNPSEAPGSIR
GDLGLTVGRNIHGSLSAEAREINLWFNENIITSVSPDAWLYE"
complement(8318. .9277)
/genes="gerCC"
complement(8318. .9277)
/genes="gerCC"
/notes="ORFID:SA1302"
/codon_start=1
/translation="hypotheical protein"
/protein_id="BAB42563.1"
/db_xref="GI:13701269"
/translation="MAKLNNEIKKVEORLEKAISKDSVLBOASLHLLSSGGKVR
PAFVLLSQSGKDEQSEQYQVAVALELHMAVLVHDDVIDKSRKGLKLTISKWD"
Query Match 6.4%; Score 119.4; DB 1; Length 301550;
Best Local Similarity 51.0%; Pred. No. 2e-16;
Matches 282; Conservative 0; Mismatches 271; Indels 0; Gaps 0;

Qy 174 AAGTGAGCAACTCGCTCAGATTAAATCTGAGCGCATGCTTTTGTCCCGAGGTC 233
Db 173176 AAGCAAAACAATCGCGAAATGTTAAATGACGGTCTTTTACCAGTTGACTTGAAAGAAA 173117

Qy 234 TCAGTGAAGACGACATCTCTTCGATCTTGGAAAAAACAATGTACACAGGCATTATCT 293
Db 173116 TTTATTCAAACTCAGTAGTGCCCAATTTGTCAGATGCAATTAGATAAACTGTGTTG 173057

Qy 294 CAGCATGCTGTGCTTGGCAAGCTTATGTTTGTATGAGCGTATATPATAGATTGGAG 353
Db 173056 CATCGTTTATTGTTGATGACATTAATTTATTATTATTCATGCTTGGATTCTACCGATTACCTG 172997

Qy 354 GCGTCATCGCTTCGGAGCTGTTCTTCGATCTTTGCTTATCTGGCGACGCTACAGT 413
Db 172996 GTTTAGTAGCGATTATTGTCATTGACATTAATCTATCTAATACGTTAGTACATTTAAT 172937

Qy 414 ATTTGGATGCGCCACTCACCTTGTGAGGACTCGCTGGGATGTTCTTCTGATGGGATGG 473
Db 172936 TCATTTCCGGGTTCTAATTTTACAGAGATTAGCGCGTGTGATTAGGTGATGATGG 172877

Qy 474 CCGTAGATGCAAAATGTTCTTGTATTCGAAAGAACCCGAGAGAAATTTTATGTCTCAA 533
Db 172876 CTGTAGATGCAAAATGTTCTTGTATTCGATGATGAGCGTATTAGGATGAATTCGATAGTAA 172817

Qy 534 GTCCTTAAATATCTAGAAAAGATATACCAAGGCTTTTGGAGCCATTTTTCATCTTA 593
Db 172816 CGATTAAGCAAGCCCTTTCTTAAAGCAACAAAGTTCATTCCTAACAAATTTTTCATCTA 172757

Qy 594 ACTTGACTACAGTATTGGCCCTCAGCAGCTTCTTTTCTTCTAGATACAGGGCTATTAAAG 653
Db 172756 ACTTAAACAAGTATTCCCGCAGCAGTATTATTCTTCTCGGTGAAGTTCAGTTAAAG 172697

Qy 654 GGTTTGTCTTGACATTTAGGAAATTTTCTCTTCAATGTTTACGGCTCTTTTTCATGA 713
Db 172696 GTTTTCGCAAGTATTATTATTAGTATCTTAATGATCTTTGTTTACAGCCGTGTTCTTAT 172637

Qy 714 CTAATTTTCTTT 726
| | | | | | | |
```

```

Db 172636 CAGATCTCTTATT 172624

RESULT 17
AP004827/c
LOCUS
DEFINITION
Staphylococcus aureus subsp. aureus MW2 DNA, complete genome,
Staphylococcus aureus subsp. aureus MW2 DNA, BCT 02-JUL-2002
AP004827 BA000033
AP004827.1 GI:21204509
SOURCE
Staphylococcus aureus subsp. aureus MW2 (strain:MW2) DNA.
Staphylococcus aureus subsp. aureus MW2
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE
1
Baba,T., Takeuchi,F., Kuroda,M., Yuzawa,H., Aoki,K., Oguchi,A.,
Nagai,Y., Iwama,N., Amano,K., Naimi,T., Kuroda,H., Cui,L.,
Yamanoto,K. and Hiramatsu,K.
Genome and virulence determinants of high virulence
community-acquired MRSA
Lancet 359 (9320), 1819-1827 (2002)
22040717
12044378
2 (bases 1 to 333750)
Aoki,K., Oguchi,A., Nagai,Y., Amano,K., Iwama,N., Baba,T.,
Kuroda,M., Hiramatsu,K. and Kikuchi,H.
Direct Submission
Submitted (06-MAR-2002) Akio Oguchi, National Institute of
Technology and Evaluation, Biotechnology Center; 2Chome 49-10
Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:oguchi@nite.go.jp, URL:http://www.bio.nite.go.jp/,
Tel:81-3-3481-8423, Fax:81-3-3481-8424)
FEATURES
1. 333750
Location/Qualifiers
/organism="Staphylococcus aureus subsp. aureus MW2"
/strain="MW2"
/db_xref="taxon:196620"
complement(278. .619)
/genes="MW1341"
complement(278. .619)
/genes="MW1341"
/notes="ORFID:MW1341"
/codon_start=1
/translation="hypotheical protein"
/protein_id="BAB95206.1"
/db_xref="GI:21204510"
/translation="MTNITKEVPDNLBQSIDVPAKNKTLGSSSEAKPYLDYHSKIIDY
FKQVNDITGNIDFDKLNQYFVPMNPKERYDYMIERYHFMGYRQMKTFKTELKMNNA
SYQTRLNKQV"
complement(624. .1283)
/genes="nth"
complement(624. .1283)
/genes="nth"
/notes="ORFID:MW1342"
/codon_start=1
/translation="endonuclease-like protein"
/protein_id="BAB95207.1"
/db_xref="GI:21204511"
/translation="MVSKKKALEMIDVIANMFPDABCELCXNDPFPFELTIAVLLSAOCT
DVLNVRVTTTELKPKYTPEDYLAWSDELMNDIRSLGYNKAKNKKLQSLIDQFN
GEIPQTHKELESAGVGRKTANVMSVAFDEPSLAVDTHVERVSKLGINRWKDNVRO
VEDRLCSVTPRODWNRSRSHQLIFFGRYHCLARKPKCDICPLEBDCREGQKRYKASLKE
A"
complement(1273. .1959)
/genes="MW1343"
complement(1273. .1959)
/genes="MW1343"
/notes="ORFID:MW1343"
hypotheical protein, similar to chromosome replication
initiation protein dnab"
/codon_start=1
gene
CDS
```

```

/transl_table=11
/protein_id="BAB95208.1"
/db_xref="GI:21204512"
/translation="MDKYLQKARPVVIRRELDLDHYSDGLGDEODLVLLKLIVASETS
NKPSILLOKSGTMOPRDIITMVIQNLIOBELLEOVQKDEGRFTYMDLPPFEKL
SHLKQOSETKEONSKKEPKOLFVREDEFPARLSPYEITLQWIDVDRKDAIIO
AALDEANSINKLSFKMDRIILNKKNNVKITIDSRKIREKFNKRNKTHVTKYTPKRD
WLNGENLDGK"
complement(2287. .3579)
/gene="asn5"
complement(2287. .3579)
/note="ORFID:MM1344"
/codon_start=1
/transl_table=11
/product="asparaginyl-tRNA synthetase"
/protein_id="BAB95209.1"
/db_xref="GI:21204513"
/translation="MKTTIKQAKDHLNODVTIGAWLTNKSQSGKIAFLQKRDGTFWQ
GVVKSVEDEVFRLAKEITQESSLYTGITTEDNRSDLGEMOVKSEIYVSEAHYR
ITPVKHGTEFLMDRHLMLRSKQHAAMKIRNEVIRATYEPFNKDGFTKVPDLITS
APEGTSELHTYKYDQDAFLQSGQLVLEAAMAGVFSFGPTFRAEKSTRHLLIE
FMWEGEMAFTHAESLEIOEYVTHVVKVLEAKCLEKILERDTSKLEKVAIFPR
ISYDATELEIKSEGFDDIEMGEDRGAPHERIAAHYDLPVPTITVPTKIKPFWQNP
ENERTVCAADLIAPEGYEITIGSERDDELLERQKHEGLDEAASYLDLRVGS
VPHCGFGLGERTVAMISGEBHRETAFFPRLNKLTP"
complement(3901. .6594)
/gene="ding"
complement(3901. .6594)
/note="ORFID:MM1345"
/codon_start=1
/transl_table=11
/product="probable ATP-dependent DNA helicase ding"
/protein_id="BAB95210.1"
/db_xref="GI:21204514"
/translation="MGMATYAVNLETTNGNDLPDDIIQIGTFVRNNOIIDRYHSMI
RTNLEIPRJIQALTSIEBNMLQAPRYNOVAOEIYDKIDCIFIPAHNVDPDLNFIKA
FKDNIQYRPKRVIDTLEIFKIAFPDTKSYQSLSELAHGGITLANAHABDEATIK
LMIAFERFEKLPDITLKQLYLSKQKYDIDFEMVROYDAKPKDKSEKATQII
YRKQVDPKPTPTNGSLKSLYSKAVQDGLTPYQDLYLAETIIDQMLESEKAMIA
SLGSGSLAYLLAALMNTIETGKHMVSTNKLQSLQLEKDI PAMNEALNFKNALL
IKESDVTI3AGLISQILKDPSTSEYNNILKQILLMTTPSGDIQELNKKGQKWF
DQKLETVPARHDVHYNFITKNAQNIQIGITNHAHLHSDVENSITQLEPDDCVDA
HRLPDVALQVNTLSYADIKYQGLIGKENEKLLAIDQLEKQILKXIDLAID
FGLKASNMKZHEHLENEOLFSTIFTIINSDVDDIHRFNHVFETPKDILKDLAID
KUNKLEIFNGISHTKTVLSKQLYLKDKNLEKAGHTSPISIKLSQKSTIR
LYKDVAVKDVLTKQVLEKEKSLIFISGLTKFNISFPAFKQLENKDHVPMNTSL
QSAKNTSVTIPSDVASYOKNIDVASYISYIETTTISSKCLVLPFSYKAMHQ
DMLNELPEFEDVVLTOQONQNYKIIVQGFNFDAIILGISTFPEGDFQANGLKCYM
IAKLRFNKNNAKWMADMSEPTSTFKGVYLPDAVTRQGGRLIRSENRGIIISFD
DRLINSYKNFPEQTELENYRQKGDIOQFGKLLRQIOKKKK"
complement(6618. .7589)
/gene="MM1346"
complement(6618. .7589)
/note="ORFID:MM1346"
/codon_start=1
/transl_table=11
/product="truncated hypothetical protein"
/protein_id="BAB95211.1"
/db_xref="GI:21204515"
/translation="MSKYSQDVQLQLLYKNKENYISGSIASESLNISRTPAVKVIDQK
LEGCKIDSVNHKGLLOQLPDIWYQGIIDQYTKSALFDFSEVYDSDISDTQAKSKL
VGNOSFFILSDQTKGRGRENRMSSSGKMGVSLVAVVANNNGIEMVIGISKNLEIAG
IRDAIQHFSODEVGVKXPNDIYIDNGYCGGLTGMVANNNGIEMVIGISKNLEIAG
NFBDSIRERATSVQLDHDKLDIYQFLERLLQIEKXKNQPTITPSEIRIEVYIAASN
IMNRTLLFTENDQPKGQALDLDYDGLVIVDEGESHRLISADID"
complement(7576. .8778)
/gene="MM1347"
complement(7576. .8778)
/note="ORFID:MM1347"

```

```

hypothetical protein, similar to poly(A) polymerase"
/codon_start=1
/transl_table=11
/protein_id="BAB95212.1"
/db_xref="GI:21204516"
/translation="MDKSLFEQARPLIEQIIONGFEAYVGSVRDYVMGNRIHDIDI
TTSATPDEIIESFSHIIPVKGEGTINVPENBENVTTFRAEEDYVDHRRPBGVTFV
RDYEDLOQRDPFMNAIAMDYKLYDYDGGQDINNRIIRTYGIAEERFOEDALMI
RCURFOSQSPDIATTFEAMRIQKADIFLSTERYVIELTKGRIANVKSFNHLS
LKAENYMPYFEHLDNMQINVTEDIDELIAIVSKEDINSKPKLSNRQYKDINO
YIQIMNALPSIITKEQKMFVYDTHLTKNVWVADYKANDIQGHEPLIVLQTDID
ETLHRLPMENRKDMVMVGGVLMAMHNAKSGPKMLDVLQIEIAIVTGKVSNEETELK
WQDNHVKI"
complement(8783. .9925)
/gene="MM1348"
complement(8783. .9925)
/note="ORFID:MM1348"
/hypothetical protein, similar to lipopolysaccharide
biosynthesis-related pr homolog"
/codon_start=1
/transl_table=11
/protein_id="BAB95213.1"
/db_xref="GI:21204517"
/translation="MKIGITCYPSMGSGGIATLGLIKLAERGHEVHTSNIPRIR
KPLPMNIFHOVENQYARQYPPYDITLSTKIAEVYKEVGLDLHMHYAVPAICGIL
ARMSGKDIKIMTTLHGTDITVLDYHSLQGAIKFIEKSDIVTSVSKLAQDTHBII
ENKKEIIPYVRENEPPTKNTALSKQFGIAPDEKVIHVSNSQYKRIIDIIEIF
AKVREKIPSKLILGDPELVPWRQLEKLANEEDVFLGKQCVSEFPQSLDVLIL
SEKESFGLTLEEMKTVGVPIGSMAGIKKVIKGGTEFVVDGDCDSASDVIRLLE
DKLYNKLQKNMADIABRGSSELIIDQYEVYYQKMLNHNKSGE"
complement(10167. .10484)
/gene="MM1349"
complement(10167. .10484)
/note="ORFID:MM1349"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB95214.1"
/db_xref="GI:21204518"
/translation="MKSVMQGEVEDEYIGQFKTGYSPLANLARLTVEYGLAREIN
HYGKKKKDSEANIKAEIQLGNLFLVLLCLANSMDIMTESFNETMEKPNTRDKRF
ERK"
complement(10820. .11518)
/gene="MM1350"
complement(10820. .11518)
/note="ORFID:MM1350"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB95215.1"
/db_xref="GI:21204519"
/translation="MSISFTMIITVPIIWLVLPIAHOHKVXSNYKXSVQSRSTGKGR
EVALBILHANGYIDVAVKGEFLDHTYPPKKVYLSPANIDRSEVGTALIAHEVG
HALQDHQGYWFLERFALVIVANLGSSLSYMTIMGLIITLALGASAGSDALMGALM
SLAVLSITVLTPVEPDASSRAMKQITALNIVEKEYHAKRVLSAAMTYVAVATAVAV
AEIVRIILLARSSD"

```

```

Query Match      6.4%; Score 119.4; DB 1; Length 333750;
Best Local Similarity 51.0%; Pred. No. 2e-16;
Matches 282; Conservative 0; Mismatches 271; Indels 0; Gaps 0;

```

```

QY 174 AAGTGAACAACTGCTCAGATTAAACCTGAGCGAGTCTTTGTCGCCAGGTTG 233
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 227352 AAGCAAAACAATGCCGAAATGTTAAATGACAGTTCTTAAACAGTTGACATGGAAGAAA 227293
NFBDSIRERATSVQLDHDKLDIYQFLERLLQIEKXKNQPTITPSEIRIEVYIAASN
IMNRTLLFTENDQPKGQALDLDYDGLVIVDEGESHRLISADID"

```

```

QY 234 TCACTGAAGAGCAGATCTCTTGATCTTGGAGAAAACAATGTACCAAGGACATTATCT 293
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 227292 TTATTTCAAACTCAGTAGGAGCCCAATTTGGTCAAGATGATTAAGATAAATCTGTGTTG 227233
CAGCATGCTGTGCTTGCGCAATGCTTATGTTTGAAGACGATATTAATAGATTTGGAG 353

```

```
Db 227232 CATCGTTATTTGGTGTAGCAATTAATTTATTTATTTATTCAGCTTGGAATCTACCGATTACCTG 227173
Qy 354 GCGTCATCGCTTCGGGAGCTGTTCTTCTCTGAATCTTTTTCGTTATCTCGGCGAGCTCTACAGT 413
Db 227172 GTTTAGTAGCGGATTTATGCAATTCGACAACTTATATCTATCTAACTAGTAGCATTTTAATT 227113
Qy 414 ATTGGATGCCCACTCACCTTGTGAGGACTCGCTGGGATGTTCTTCTGCTATGGGATGG 473
Db 227112 TCATTTCCGGGGTTCTAACTTTACAGGATTAGCGGGTGGTATTTAGGTGTAGGTATGG 227053
Qy 474 CCGTAGATGCAATGTTCTTGTATTCGAAAGATCCGAGAGGAATTTTATTTGCTCAA 533
Db 227052 CTGTAGATGCCAATTTATCATGTATGAGCGTATTAAGGATGAATCTCGAATAGGTAGAA 226993
Qy 534 GTCTTAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGAATCTA 593
Db 226992 CGATAAGCAAGCCTTTCTTAAAGCAACAAAGTTCATTCCTCAACAATTTTGAATCTA 226933
Qy 594 ACTTGACTACAGTATGGCTGCTAGCACTCTTTTCTTCTCTAGATACAGGGCTATTAAG 653
Db 226932 ACTTAACAACAGTATATCGCGCAGCAGTATTTCTTCTTCTCGGTGAAAGTTCAGTTAAAG 226873
Qy 654 GGTTCGTTTGGACATGATTTTAGGAATTTCTCTTCAAGTTTACGGCTTTTTCATGA 713
Db 226872 GTTTCGGCAAGATTTATTTATAGGTATTTCTATGATCTTTGTACAGCCGTGTTCTTAT 226813
Qy 714 CTAAATTTTCTT 726
Db 226812 CAAGATCTTATT 226800
```

```
RESULT 18
AP003363/c
LOCUS
DEFINITION
Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
sequence, section 6/9.
ACCESSION
AP003363 BA000017
VERSION
AP003363.2 GI:14247399
KEYWORDS
SOURCE
Staphylococcus aureus subsp. aureus Mu50 (sub_species:aureus Mu50,
strain:Mu50) DNA.
ORGANISM
Staphylococcus aureus subsp. aureus Mu50
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE
1
AUTHORS
Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,
Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Lian,J., Ito,T., Kanamori,M.,
Matsumaru,H., Naruyama,A., Murakami,H., Hosoyama,A., Kaito,C.,
Mizutani-U.I., Takahashi,N.K., Sawano,T., Inoue,R., Kaito,C.,
Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S., Yabuzaki,J.,
Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C.,
Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiramatsu,K.
Whole genome sequencing of methicillin-resistant Staphylococcus
aureus
Lancet 357 (9264), 1225-1240 (2001)
JOURNAL
MEDLINE
21311952
REFERENCE
2 (bases 1 to 342600)
AUTHORS
Ohta,T.
TITLE
Direct Submission
JOURNAL
Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College
of Medical Technology and Nursing, Department of Medical
Technology; 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan
(E-mail:tohta@akura.cc.tsukuba.ac.jp, Tel:81-298-53-3454,
Fax:81-298-53-3454)
COMMENT
On May 29, 2001 this sequence version replaced gi:13875943.
FEATURES
Location/Qualifiers
source
1..342600
/organism="Staphylococcus aureus subsp. aureus Mu50"
/strain="Mu50"
/sub_species="aureus Mu50"
/db_xref="taxon:158878"
complement(42..815)
/gene="SAV1628"
```

```
CDS
complement(42..815)
/gene="SAV1628"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57790.1"
/db_xref="GI:14247400"
/translation="MKHOFSRNELAIGQGLDLLKQQTIVILGVGVGFAAEALART
NIGHTILDKDDVIDTNVNRQHALTTVQSKVSLMEERIKLINPDCKVTPLHMYFT
BETYEIEDNYDIDYFVSDADTIRMYKVLMEKCLDRGKIVISSMGAANTDPTFTFIS
DKSTHTPMAKIIRQKLKGLIRKGIIPVPSDESPIVIREDVKDIVGDKNAINRKGQ
MPPSNAPVPSVVGILICASYVNDVKDIPVRRIKDKG"
complement(1029..1130)
/gene="SAV1629"
complement(1029..1130)
/gene="SAV1629"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57791.1"
/db_xref="GI:14247401"
/translation="MGFLAAATPYMKDVQNKQKGTHLYITGRMIKLL"
complement(1276..3042)
/gene="asps"
complement(1276..3042)
/gene="asps"
complement(1276..3042)
/gene="SAV1630"
complement(1276..3042)
/gene="SAV1630"
/codon_start=1
/transl_table=11
/product="aspartyl-tRNA synthetase"
/protein_id="BAB57792.1"
/db_xref="GI:14247402"
/translation="MSKRTTYCGLVTEAFLGQEIITLKGWNNRRDGLGILFVDLRDRE
GIVQVFNPAFSEALKIAETYSVVVEVQVTVKRDPETVNPKIKTQGVQVQVNI
KIVKSETPPSPSINEENVNVDENIRKYRLDLRQELAQTFKMRHQITRSROYLDD
EGFPDIEPVLTKSTPEGARDYLVSFRHDEGFALPQSPQLFKOLLMSGDKYQI
VKCFREDLRADQPEFTQVDIEMSFVDQEDVMQGEEMLKVKVEKVGVEINGATPR
MTYKAMRRYSGDKPTPEMELIDVSQLRDMDFKVDKTVENDEGKAIKIVAKGAAE
QYTRKMDALTEFNIYGAKLAVKVKVVEDGLTGPGRFETENVTETLLTGAEGAD
LVMPVADKPNVVAQSLGARVKLAKELIDETKLNFLWVTOWPLLEYDEDAKRYVAA
HHPTVSPKEDIKLGTPAEAEANAYDIVLNGVELGGGIRIHGSELOEKMPFVLGF
TKEAQEQFGFLDDAFKVGAPPHGGIALGLDLRLMLLTNRNLRTIAPKPTASATCL
LTNAPGEVSDKQESLSLRIRH"
complement(3058..4320)
/gene="hiss"
complement(3058..4320)
/gene="hiss"
complement(3058..4320)
/gene="SAV1631"
/codon_start=1
/transl_table=11
/product="histidyl-tRNA synthetase"
/protein_id="BAB57793.1"
/db_xref="GI:14247403"
/translation="MIKIPRGQDILPDSKQKRYIENOLDMLTFTYNYKEIRTPIFE
STDLFARGVGDSTDVQKEMYTFKDKGRSITLRPEGTAAVRSYIEHKMQGNPQPI
KLYNGPMFPRYERKQKQRFQNFQVEAIGNPSVDAVLAMVNHIIYOSFGLKHL
LVINGVDMASKEYNEALVHFEPVIEHFCDCQSLRHPNMRILDCVKDRDKEAIK
TAPRITDPLNEESKAYEVOKAVLDLGIPIYEDNLVRGLDYTHTAPELMDNPNY
DGAITTLCCGGRYNGLLELDGSPETGIFALSIERLLALSEEGELDIEENLDLPI
VMTGQADRYAVKLNHLNHNKIGKADKYLQRKIKGQMKQADRUGAKFTTIVIGDQBLE
NNKIDKNNMTTGESETIELDALVEYFKK"
complement(4782..5657)
/gene="lytH"
complement(4782..5657)
/gene="lytH"
complement(4782..5657)
/gene="SAV1632"
/codon_start=1
/transl_table=11
/product="N-acetylmuramoyl-L-alanine amidase"
/protein_id="BAB57794.1"
/db_xref="GI:14247404"
/translation="MKKEAWLSKKGLKNKRTLIVVIAFVLFIPLFLLNSNSEDG
NITTENALRTGPNAAPVVIYKVEGDHFKKIGKVGKWIIVEDTSSNEKGIAGWHT
```



```

NDIADNTKEKNPLOGKTYLIDPHGGSDGASNTKYSLEKDYTLTKAKELORTL
EKSGATYKMRITDPTTYSLERNIDKGAAYLSIHNDALSSNAAMTYVYHNDORALA
DTLDATTOKKGLSNKGRSDENYQVLQTKPAVLELGIISNDETMTIKDQJHRO
LBOAIVDGLKIFYSA"
/gene="SAVI633"
/complement(5654. .6106)
/gene="SAVI633"
/complement(5654. .6106)
/gene="SAVI633"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="BAB57795.1"
/db_xref="GI:14247405"
/translation="MKVYQVRKASVTNPTLNNQIKKCYCLAVIGIGONSPEQDAVY
AKTKANARLFEDDNNKLNFNITQQMNGEILSYQFTLADYKGNRPFSKSNPDQAV
KIYEFNDALRAYGLTVKTGEFTHMNVISINDGPVITLYESQDKIQ"
/complement(6118. .8307)
/gene="re1a"
/complement(6118. .8307)
/gene="re1a"
/note="SAVI634"
/note="SAVI634"
/codon_start=1
/transl_table=1
/product="GTP pyrophosphokinase"
/protein_id="BAB57796.1"
/db_xref="GI:14247406"
/translation="WNNVEPYSADEVILHAKASYLSADEYEVYLSHYIAYBAHQGR
KNGLPYIMHPLOVAGILTEMRLDGPITVAGLHVIDPTPEDEVKEMNEEVARIV
DGVTKLKKVYRSKEEOQAEHNRKLFIAADVAVIIVKLADRLHNRRLKAMREQ
IRISRETLHAYAPLALHGLITKMELEDPTLRVIDVQYERIVNLKMKSESEAYI
ETADIRTEMEDRNINBGIDNGRPKHYSITRKMKKKKQKDDOIFDLAIRVINSNE
DCAIILGHLTKMPMPGRPKDYIANKONLYSHLTHVVGPNDDPELQIRTFDME
IAEYVAHMAHMYKEGKAYSEKQTYQNKMLKSLAABDHSQDAQFMEITLDLOS
DKYAFTPASDVIELPYGAVIDPAVAIHSEVGNKMGAKVNGKIVPIDYLQTDIV
EIRTSKSYGSPRDMKIVKSSSAKGIKSPFKQDSQISQIKLETERQILDKOPALR
VEDILTEKQIQVNEKYNFANEDLPAVAGGVTSQIYVKLTERQILDKOPALR
BOEYTKSLPIKDNITDSGYVEGLENVLIKLSKCPNIPEDDITVGITGHGKAYR
TDCPNITKETERLINEMVAKSKATQCYQVLDERTAYDRKGLNEVQANVSITAGNLI
KVSGRSDIDKKAIIINISVMKANDVTRVVEKIKQLDGVYTVTRVMN"
/complement(8735. .9253)
/gene="apt"
/complement(8735. .9253)
/gene="apt"
/note="SAVI635"
/note="SAVI635"
/codon_start=1
/transl_table=1
/product="adenine phosphoribosyl transferase"
/protein_id="BAB57797.1"
/db_xref="GI:14247407"
/translation="MDLKQYSEVQDMPKRVSPKDTITIMDNSEAYATDKIVEYA
KDRVDIVVGEARGFTIGCPVATSWGIFAPVRKEGLPREVIRIRYEDLSEGTNVL
MKDAIKPQGVILITDLDLATGTIEAIAIKVEKLGIVGIAPIELKYNLGEIKIK
DYVMSLSIYDE"
/complement(9275. .11548)
/gene="SAVI636"
/complement(9275. .11548)
/gene="SAVI636"
/note="SAVI636"
/note="similar to single-strand DNA-specific exonuclease"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAB57798.1"
/db_xref="GI:14247408"
/translation="MIKPKYKMLTKPAEYISDELTSKLKLPVKKILBSKSIIDQ
AIESIISTDVNDALDSDMTKALERIKRAINDEKLIYGDYDADGVSTTLVVT
LQILGAVGMHIPIRFTGEGVGNELAPRNADDEGTLITVDNIOGHNETKMWODG
VDVITVDHEIGSTLPKAYAVIHPMHSPVPRPOOLGAGAYVLAQALLENVDYK
ALVAIGTADLVSLIDENRSLVQGLKYNQCTSVKALKEAGINDIDEFTGTI
IGPRLNAVGRDASLACELMTDDEBAALAOQVHFNNERKDIYAATTEEMANA
ETKVKGCLFLALKEMWHEBVLGIVASKIVETALPTLILINIREONHAGSARYD
QVSMEEIISAHOELIAKFGGHMAAGMTMDIENTESLAEGNKKMKELSKTSLDPYK
PVDVLTENDITIKNIDMNLRPFGTDFSRPIFEMDLSVSSVKAIQOONHKLTL
GESNIAALFMONGHLEPBLQDQPIINIGSVQINENKNGNSPOLIIDIANEQIID

```

```

YRSKRSLPFTENDENIYVLIHPKSDPVANEEYVEEIKKOOTDKVVRDLPTSMEDL
SNELQOLOSQATYIVLOHNSIYFPDGIIPMDIPEKKYKXILITQENINIOKESGLLOH
LSKVPDTLKMFLKFDLKFVIOEDGLIRINQOPDRKSIDSKVYOLRQQRMDVERQL
LYDFSEIKNWKISQLS"
/complement(11751. .14030)
/gene="secf"
/complement(11751. .14030)
/gene="secf"
/note="SAVI637"
/note="SAVI637"
/codon_start=1
/transl_table=1
/product="protein-export membrane protein SecF"
/protein_id="BAB57799.1"
/db_xref="GI:14247409"
/translation="MKSSRIIAFLLLVLLFRAGMAATYKVIKVNGLDLOGGFV

Query Match      6.4%; Score 119.4; DB 1; Length 34260;
Best Local Similarity 51.0%; Pred. No. 2e-16;
Matches 282; Conservative 0; Mismatches 271; Indels 0; Gaps 0;

QY 174 AAGTAGCAAACTGCTCCAGATTAAATCTGGAGCATGCTTTGTCGAGGATTC 233
DB 13357 AAGCAAAACAATCGCGAATGTGTTAAATGACAGTTCTTTAACGACTTGAAAGAAA 13298
QY 234 TCAGTGAAGAGACATCTCTTCGATCTGTGGAAAAACAATGACACAGGATTANCT 293
DB 13297 TTTATTCAACTCGTGTGTCGCCAATTTGTCAGATGCAATTAAGATAAACTGTGTTG 13338
QY 294 CAGCATCGTGTGCTGGCAATGCTTTATGTTGATGACGATATTAATAGATTGGAG 353
DB 13227 CATGTTATATGGTGTGACATTAATTAATTAATCATGCTTGATCTCACGATTACCTG 13178
QY 354 GCGTCATCGCTCGGAGAGCTGTTCTTGAATCTTTTGCCTTATGTGGCGAGCTTAAGT 413
DB 13177 GTTTAGTAGCATTAATGACATTAACAATTATCTTAACGTTAGTACATTATAT 13118
QY 414 ATTTGATGGCGCACTGACCTGTGACAGCTCGGTGGATTTGTTCTGATAGGGAGTGG 473
DB 13117 TCAATTCGGGGGTTCTTAACCTTTACAGAAATTAGCGGGGTTGGTATAGGTAGTATGG 13058
QY 474 CCGTAGATGCAATATGTTCTGTATTCGAAAGATCCGAGAGATTTTATTTGTCTCAA 533
DB 13057 CTGTAGATGCAATATTAATCATGATAGAGGTATTAAGATGAATCTCGAATAGTAGAA 12998
QY 534 GTCTTAAAAAATCTGTGAAAAAGAAATACCAAGCTTTTGAAGCATTTTATTTCTA 593
DB 12997 CGATTAAGCAAGCTTTCTTAAGCAAAACAAGTTCAATCTTAACAATTTTGAATCTA 12938
QY 594 ACTTGATCAAGATTTGAGCTGACGACCTGCTTTCTTCTGATAGACAGGGCTATTAAG 653
DB 12937 ACTTAACAACGTTATGCGCGACAGATTAATTTCTTGGTGAAAAGTTCAAGTTAAG 12878
QY 654 GGTTCCTTTGACATGATTTTAAAGATTTTCTTCAATGTTTACCGCTCTTTTCAATGA 713
DB 12877 GTTTCGAGACATGTTATTAATTAAGTATTTCAATGATCTTTGTTAGACCGGTGTTAT 12818
QY 714 CTAAATTTTCTT 726
DB 12817 CAAATTTCTTAT 12805

RESULT 19
AX067447/c 45613 bp DNA linear PAT 24-JAN-2001
LOCUS AX067447
DEFINITION Sequence 22 from Patent WO0078968.
ACCESSION AX067447
VERSION AX067447.1 GI:12545067
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 45613)

```



AUTHORS Lagace,R.E., Patterson,C. and Berg,K.L.  
TITLE Nucleotide sequences of moraxella catarrhalis genome  
JOURNAL Patent: WO 0078968-A 22 28-DEC-2000;  
REFERENCE Incyte Genomics, Inc. (US)

FEATURES Location/Qualifiers  
source 1. 45613  
/organism="Moraxella catarrhalis"  
/db\_xref="taxon:480"  
BASE COUNT 13249 a 9346 c 9947 g 13071 t  
ORIGIN

Query Match 6 3%; Score 118; DB 6; Length 45613;  
Best Local Similarity 50.5%; Pred. No. 5.1e-16;  
Matches 286; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

Qy 150 TCTCAGGAAATTTACCCACCGTGAAGTGAGCAAACTCGCCTCAGATTTAAATCTGCAG 209  
|||  
Db 28889 TCACAGGCTTAGATTCAGATGCAGAGCAGGAGCTTGCCTTACTGTTACGCTCAGGCG 28880  
|||  
Qy 210 CGATGCTCTTTTCCCGAGGTTCTCAGTGAAGAGACGATCTCTCTGATCTTGGGAAAA 269  
|||  
Db 28829 CGTTAGCGGCACCGATGATCTTGTGAAGAGCGAACCATTTGGCCCATCTTTGGGTCAAG 28770  
|||  
Qy 270 AACATGTACACAGCGATATCTCAGCATCTGTGGCTTGCATGCTTATGTTTGA 329  
|||  
Db 28769 ATAATATCGATAGAGATTTCTCTACACAGGTTGGCTATCTGTTGTTATTTATGGA 28710  
|||  
Qy 330 TGAGCGGTATATTATAGATTGGAGCGCTCATCGCTTCGGGAGCTCTCTCTGATCTTT 389  
|||  
Db 28709 TGATTGCTTTTATCGTCAATTTGTTGATGCAAAATTTGCTTTAGCCCTTAATGTCA 28650  
|||  
Qy 330 TCGTTATCTGGGACGCTTACAGTATTTGGATGCGCCACTCACCTTGTGAGACTCGCTG 449  
|||  
Db 28649 TTATCATGTTGCTCAATTTATGCTATTTAGGATCATCGCTCACCTACCTGCTATGCTG 28590  
|||  
Qy 450 GGATTGTTCTTCTATGGGATGCGCTAGATGCAATGTTCTTGATTCGAAAGATCC 509  
|||  
Db 28589 GGATTGTTGACCAATGGTATGGCGGTTGATGCCAATGTGCTAAATTTTGGCGGTATTC 28530  
|||  
Qy 510 GAGAGGAATTTTATTTGCTCAAGCTTTAAAAATCTGTAGAAAAGGATATACCAAGG 569  
|||  
Db 28529 GTGAAGAGCTTCCCAATGCTACTCGTCANAAATCGCCATTTGTGCGAGTTTGTATCGTG 28470  
|||  
Qy 570 CTTTGTGAGCAATTTTGTATCTAATCTGACTACAGTATGAGCTTCAGCACTCTTTCT 629  
|||  
Db 28469 CATTTAGCTCAATTTTGTATGCAATATACCACTATTAATTCATTTATTTATTTG 28410  
|||  
Qy 630 TCCTAGATACAGGCGCTATTAAGGGTTTGTCTTGACATGATTTTAGCAATTTCTCTT 689  
|||  
Db 28409 CCATCGGAACAGGACCAATCAAGGTTTGTCAATCAGCTTGCATTTGCTATTCAGTT 28350  
|||  
Qy 690 CAATGTTTACGCTCTTTTTCATGACT 715  
|||  
Db 28349 CTTTGTACTGCGATTTTAGTCACT 28324  
|||

RESULT 20  
AE013703 10963 bp DNA linear BCT 26-JUL-2002  
LOCUS AE013703  
DEFINITION Yersinia pestis KIM section 103 of 415 of the complete genome.  
ACCESSION AE013703 AE009952  
VERSION AE013703.1 GI:21957732  
KEYWORDS  
SOURCE Yersinia pestis KIM.  
ORGANISM Yersinia pestis KIM.  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Yersinia.

REFERENCE 1 (bases 1 to 10963)  
AUTHORS Deng,W., Burland,V., Plunkett,G. III, Boutin,A., Mayhew,G.F.,  
Liss,P., Perna,N.T., Rose,D.J., Mau,B., Zhou,S., Schwartz,D.C.,  
Fetherston,J.D., Lindler,L.E., Brubaker,R.R., Plana,G.V.,  
Straley,S.C., McDonough,K.A., Nilles,M.L., Matson,J.S.,  
Blattner,F.R. and Perry,R.D.

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

Genome Sequence of Yersinia pestis KIM  
J. Bacteriol. 184 (16), 4601-4611 (2002)  
12142430  
2 (bases 1 to 10963)

Deng W., Burland V., Plunkett, G. III, Boutin, A., Mayhew, G.F.,  
Liss, P., Perna, N.T., Rose, D.J., Mau, B., Zhou, S., Schwartz, D.C.,  
Fetherston, J.D., Lindler, L.E., Brubaker, R.R., Plana, G.V.,  
Straley, S.C., McDonough, K.A., Nilles, M.L., Matson, J.S.,  
Blattner, F.R. and Perry, R.D.

Direct Submission  
Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445  
Henry Mall, Madison, WI 53706, USA

FEATURES  
source

Location/Qualifiers  
1. 10963  
/organism="Yersinia pestis KIM"  
/strain="KIM"  
/db\_xref="taxon:187410"  
119..1189  
/genes="queA"  
/note="y0990"  
119..1189  
/genes="queA"  
/function="enzyme; aminoacyl tRNA synthetases, tRNA  
modification"  
/notes="synthesis of queuine in tRNA; residues 1 to 354 of  
356 are 83.33 pct identical to residues 1 to 354 of 356  
from E. coli K12 : B0405; residues 1 to 353 of 356 are  
83.56 pct identical to residues 1 to 353 of 354 from  
GenPept : >gb|AA119358.1| (AE008714)  
S-adenosylmethionine-tRNA ribosyltransferase-isomerase  
[Salmonella typhimurium Lt2]"  
/codon\_start=1  
/transl\_table=11  
/product="probably S-adenosylmethionine:tRNA  
ribosyltransferase-isomerase"  
/protein\_id="AA084571.1"  
/db\_xref="GI:21957733"  
/translations="MRVADFSELPALIAHPQQRSGCRLLSLDGTGTLTHGIFT  
DLIDKLAPGLLVNTRVIFARLFGKASGGKLEVLVLDHHRVLAHVKAASKAPK  
PGALLGDDDESIRATMLARHDTLFLRFDERDVFILNAVGHMPLPPYIDRDEDA  
DRELYTVYSORPCAVAAPTAGLHDEPMALAOEKGLEMAFVTLHVAGCTFOPVRVD  
TIEDHMHSEYAEVPOEVDAVLACARGKRVAVGTTSVRSLESAKAAENGLIAPF  
FGTRTFIYVGHYQVVDALVTNPHLPSTLIMLSAFAGYKNTMNAQQAVABQYRP  
FSYGDAMPISRNPRAPQEKVSP"  
1282..2406  
/genes="tgt"  
/note="y0991"  
1282..2406  
/genes="tgt"  
/function="enzyme; aminoacyl tRNA synthetases, tRNA  
modification"  
/notes="residues 1 to 374 of 374 are 89.57 pct identical to  
residues 1 to 374 of 375 from E. coli K12 : B0406;  
residues 1 to 374 of 374 are 89.83 pct identical to  
residues 1 to 374 of 375 from GenPept : >gb|AA119359.1|  
(AE008714) tRNA-guanine transglycosylase [Salmonella  
typhimurium Lt2]"  
/codon\_start=1  
/transl\_table=11  
/product="tRNA-guanine transglycosylase"  
/protein\_id="AA084572.1"  
/db\_xref="GI:21957734"  
/translations="MKVELQKTDGRARRGRLVFERGVVETPAPMPVCTGTVKGMTPE  
EVKETGAQLLGNTHLMWRPGQEIIMKLHGLDHFQMHGPIILTDGSGFQVSIKANR  
KIXEKGVHKFNINGSVFLSPKSEHIEQYDLGSDIYMFDECTPPFADMDYAKRME  
MSLRWAARQRFDLNNALFGIITGGVIEDLRDVSVKGLVDIGVAVGGLAVG  
LPTDGVVKIRNAKHKDDTATLDEHCDCTCRHYSRAYLHLDRCNEILGARLNTIHN  
LRYVQRLMAGLRQAIIEGKLEHFVEDFYGRIGKVPVPLNV"  
2452..2853  
/genes="y0992"  
2452..2853  
/genes="y0992"

gene

CDS

gene

CDS

gene

CDS



```
Db 4526 CTTCTCAAGTATCGTCGATGCGGAACAATAACGACGCTTATCACGCCATTATCTTTATG 4585
Qy 630 TCCTAGATACAGGGCCTATTAAAGGGTTTGTCTTGACATGATTTAGGAATTTTCTCTT 689
Db 4586 CTGTGGTACTGTTTCGATTAAAGGCTTGTCTATTACACGGCAATTGTTGTAACT 4645
Qy 690 CAATGTTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGGATGAATGAACCC 749
Db 4646 CAATGTTTACGGCAATAGTCGGTATCTCGTGCCATCGTCAACCTGCTTACGGCGGCAAGC 4705
Qy 750 AACATACACAGTGCATATGATGATAGTATGTCGGGATAAAGCATGATTTCTTGAGAG 809
Db 4706 GCATTAAACAGCTGTCTAT--TTGAGGAGTACGTTGTGGCAGAGATTATCTGTTGAAC 4763
Qy 810 GATGCAAAAAAATTTTGGCTGTTCTGGAAGTGTCTTTTCTTTAGTTGCGTTGCTCTCG 869
Db 4764 AACTAAACTATGGCGGTAAGTCTATGACT--TTATGGCTGGGATACGTTGCTTCG 4820
Qy 870 GGTTTGGAGCCTGGAAATTCGGTTTGGGAATGGAATTTAAAGAGGGGTATGCTTTACCT 929
Db 4821 CGGTTTCACTGTTGTTGTAGTCGCATCCATTGTGGTGATGTCGACTAAGGGATTCAACT 4880
Qy 930 TTAATCCAAAGAGCATGGCATCAGCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 989
Db 4881 GGGGTCTGGATTTTACCGGTGGTACGGTGATGAAATTAATCTGGAAACCCCTCGCGAT- 4939
Qy 990 AACTACAGGAAGCTGTTCTTCTCTAGAGACTTCGGTATTCAACATTTGATCTTCAG 1049
Db 4940 -----CTTGACCAACTCGTGACAGCTGGTGAACCGGGTTTGAATCGCGG 4987
Qy 1050 AAAAGATCAAAATCTATTATTAGTATAAGCTTTTAAGCTATATACTAAGCAGATACGAGCCT 1109
Db 4988 ATACTGCAAAACTTTGGT-----AGCTCTCGTATGTCATGTCGATGTCGCCACC 5037
Qy 1110 CTCCTCTAAATTAACGATCATGATGCTGGCTTATTTGGGATTTGTCGAGAAACAGGC 1169
Db 5038 TGCAACCGGCACTGCGGCGGAGGAGCTGG-GTAACAAAATCATCTCGGTTATCAATGAAT 5096
Qy 1170 CTAGATTTCTCTACGGAACCTTAACGAAACGCAAAATTTTGGTCAAGGTAAGCAGCA 1229
Db 5097 CTGTTGATAGATGATCAGTAAGCGATATGAAATTCGTTGGGCTAGTGTGGCAGTG 5156
Qy 1230 AACTATCGAAGAAATGCGTTATACGGCGACCATCGGCTTTTAGGAGCTTTGGCAATCA 1289
Db 5157 ATTTAGC-----CCAAAGCGGTGCTCTGGCGCTTTTGTCTGCGTTACTGTCTA 5204
Qy 1290 TCTTGCTCTATGTAGTTTGGCTTTGAATGGCAATATGCTTTAGTCGCGTATGCGCTT 1349
Db 5205 TTCTGGTTTACGTGGGTTTCCGTTTGAATGGCGTCTGGCATTTGGGGCGGTTATCTCAT 5264
Qy 1350 TAATTCATGACCTTTTGGCTACCTGTGAGTCTTCTTTTATAGCATATTCCTTTTGAAGA 1409
Db 5265 TAGCCCATGAGTGTCTATTACCATGGGAGTCTTG-----TCGTTGTTC 5309
Qy 1410 AAATTCAAATAGATTTGAAGCATTGTTGCTTTAATGACTGTATTTGGGGTATTCATTAA 1469
Db 5310 ATATTGAGATTTGACCTGACGATCATTTGCTTCATTGATGTCGTTATTGTTACTCTCTGA 5369
Qy 1470 ACAATACTTTGATCATTTTGTGATGTTATGTTGAAGATGCGGCAAGCAACCTGTTTACC 1529
Db 5370 ACGATAGTATCGTGTATCAGACCGCTATTCTGTGAGAACTTCGCGCAAAATTCGTCGTGAA 5429
Qy 1530 CTATGCATGTTTATGATGATGATGCTTCAAAAGACGCTTCAGCGCGCATGATGACAA 1589
Db 5430 CGCCTTAGCAAAATCAAGATGATCTCTGACACAAACCTTAAGTCGTACGATTGACTT 5489
Qy 1590 CAGCTACAACCTCTATCAGTTTTTGTAAATGCTTTTCTTTTATAGCGCGCTCTCTGCTTTA 1649
Db 5490 CTGCGACGACTCTGATGTTGTTGTATGCTGCTTTATCTTCTGGTGTGGATGCTGCAAG 5549
Qy 1650 ATTTTGCATTTATTATGACCATAGGATTTCTTCTAGGAACCTTTATCGTCTCTTTATATTG 1709
```

```
Db 5550 GCTTCTCATTAACGATCGTATCGGTGTTACTACTCGGTACCGTTCATCTATCTATCAGTGG 5609
Qy 1710 CACCACCTCTGTTGTTG 1726
Db 5610 CGTCCGCGCTGGCGCTG 5626

RESULT 21
AJ414155/c
LOCUS
DEFINITION
  Yersinia pestis strain CO92 complete genome; segment 15/20.
ACCESSION
  AJ414155 AL590842
VERSION
  AJ414155.1 GI:15980975
KEYWORDS
  Yersinia pestis.
SOURCE
  Yersinia pestis.
ORGANISM
  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
  Yersinia.
REFERENCE
  1 (bases 1 to 205050)
  Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G.,
  Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L.,
  Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdano-Tarraga, A.M.,
  Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G.,
  Feltwell, T., Hamlin, N., Holtroyd, S., Jagels, K., Leather, S.,
  Karlyshev, A.V., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K.,
  Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrall, B.G.
  Genome sequence of Yersinia pestis, the causative agent of plague
  Nature 413 (6855), 523-527 (2001)
  21470413
  MEDLINE
  11586360
REFERENCE
  2 (bases 1 to 205050)
  Parkhill, J.
  Direct Submission
  Submitted (04-OCT-2001) Submitted on behalf of the Yersinia
  sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
  Hinxton, Cambridge CB10 1SA E-mail: parkhille@sanger.ac.uk
  Notes:
  Details of Y. pestis sequencing at the Sanger Centre are available
  on the World Wide Web.
  (URL, http://www.sanger.ac.uk/Projects/Y_pestis/).

FEATURES
  source
    location/Qualifiers
    1..205050
    /organism="Yersinia pestis"
    /strain="CO92"
    /db_xref="taxon:632"
    /note="biovar: Orientalis"
    69..1197
    /gene="YPO3021"
    /pseudo
    69..1197
    /gene="YPO3021"
    /note="Similar (after the first 34 N-terminal aa) to the
    previously sequenced Yersinia pestis outer membrane
    protein YopM or Yop48 SW:YOPM_YERPE (P17778) (367 aa)
    fasta scores: E(): 3.9e-25, 37.5% id in 339 aa, and to
    Yersinia enterocolitica YopM protein TR:P74988
    (EMBL:AF102990) (367 aa) fasta scores: E(): 8.2e-26, 37.2%
    id in 339 aa. There is a possible frameshift following
    codon 188. The frameshift occurs at a homopolymeric tract
    of 8C. The sequence has been checked and is believed to be
    correct"
    /pseudo
    /codon_start=1
    /transl_table=11
    /product="putative outer membrane protein (pseudogene)"
    258..326
    /gene="YPO3021"
    /note="Pfam match to entry PF00560 LRR, Leucine Rich
    Repeat, score 14.40, E-value 2.8"
    /pseudo
    378..443
    /gene="YPO3021"
    /note="Pfam match to entry PF00560 LRR, Leucine Rich
```

```

misc_feature
Repeat, score 16.20, E-value 0.79"
/pseudo
447. .512
/genes="YPO3021"
/notes="Pfam match to entry PF00560 LRR, leucine Rich
Repeat, score 16.30, E-value 0.71"
/pseudo
516. .561
/genes="YPO3021"
/notes="Pfam match to entry PF00560 LRR, leucine Rich
Repeat, score 17.80, E-value 0.26"
/pseudo
1403. .2104
/genes="YPO3023"
1403. .2104
/genes="YPO3023"
/notes="Similar to Escherichia coli hypothetical 24.1 kDa
protein in glf-nant intergenic region YhcJ SW:YHCJ_ECOLI
(P45426) (229 aa) fasta scores: E(): 0, 59.7% id in 226
aa, and to Pasteurella multocida conserved hypothetical
protein YhcJ TR:Q9JL64 (EMBL:AF237924) (228 aa) fasta
scores: E(): 0, 58.2% id in 225 aa. It is also similar to
the N-terminal region of Streptomyces thermoviolaceus
beta-N-acetylglucosaminidase precursor NsgA TR:O82840
(EMBL:AB08771) (632 aa) fasta scores: E(): 0.86, 25.1% id
in 207 aa"
/codon start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAC92265.1"
/db_xref="GI:15980976"
/db_xref="SPTRMBL:Q8ZCG9"
/translacion="MSNLNLRLKQLONGLIASCPVPSGMSMDPEIYAAMACALAGG
AGLRREGISIOAVRATDAPIIGIRKDPDSERVITPWLIEDILASAGADIIF
DYTCGERPSVADLYORARATGCLTMADASINIDGLAHLHGIDFIGTIGSGYQAVT
PREPDLATVQALQAGCRVIAGRHSFALMAAASAGAVYVTSAGATIREHICGNF
CDALKQCEATKLEY"
167. .3060
/genes="YPO3024"
2167. .3060
/notes="nana; np1"
/genes="YPO3024"
/EC_number="4.1.3.3"
/notes="Similar to Escherichia coli N-acetylneuraminate
lyase subunit Nana or Np1 SW:NP1_ECOLI (P06995) (296 aa)
fasta scores: E(): 1.7e-16, 29.1% id in 265 aa, and to
Vibrio cholerae N-acetylneuraminate lyase, putative VC1776
TR:Q9KRG6 (EMBL:AE004255) (298 aa) fasta scores: E(): 0,
75.1% id in 293 aa"
/codon start=1
/transl_table=1
/product="probable N-acetylneuraminate lyase"
/protein_id="CAC92266.1"
/db_xref="GI:15980977"
/db_xref="SPTRMBL:Q8ZCG7"
/translacion="MKKLTGLIAPHTFFDGEVNYNPEVIDOIAEHLINDGVKGVYVC
GTTGGEVSGYDERKILAEKRWYNAAGKLSITLIGALSIRDADLSRHMETDIPFPT
SAIIGPHEFPGNLIDLINCAALIAAASKGYVYHSGMSGVNIDMEOFILKASKIP
NLSGIKENNADLYEFQRCIRVSGSGKFTDIPGVDEHLPGSLAVGATGVNVAAPLR
FKHITADFNAGQVAVORGMDFVIALIRLVLFSGVAAAGKAMQLHIGDAENPRLR
AATKEOKQVYVRNMDAITLQ"
2236. .3300
/genes="YPO3024"
/notes="Pfam match to entry PF00701 DHDPs.
Dihydrodipicolinate synthetase family, score 87.30,
E-value 3.1e-22"
2284. .2337
/genes="YPO3024"
/notes="P500665 Dihydrodipicolinate synthetase signature
1."
complement(3158. .4057)
/genes="YPO3025"
complement(3158. .4057)

```

[illegible]



goal is to make the genome sequence more useful. Comments to the authors are appreciated.

```

FEATURES
    source
        location/Qualifiers
            1..10899
                /organism="Rickettsia conorii"
                /strain="Malish 7"
                /db_xref="taxon:781"
                complement(103..131)
                /note="REP01, repeated element"
    repeat_region
        gene
            154..1098
                /gene="RC0885"
            154..1098
                /gene="RC0885"
            /codon_start=1
            /transl_table=11
            /product="unknown"
            /protein_id="AAL03423.1"
            /db_xref="GI:15619991"
            /translation="MTGLPSRLILAMTISIDSSOYOISRTNMLIKLKHIPVDY
            DVFLPDMGVTIGGCTTGGVONINKITIRKAVYTYTNAPRNIFSLHOTIKSGINA
            EPGMIISSGKIIVEMILESKEKPGIEKPAIYHGLNDINISIQPTIDDKANKAF
            LMTIYRDENLNLNEFENELFKIVQKRVNINCANPLGINOHGYRVKSGYAEKIK
            QLGCKVYSGKPYEIEYSKIKECHNTPKRMIMIGDVTYTDILANRLGIDBSALVLT
            GNSREXHIEFDNIDEXKIDSLMKASVKOSITPSPFVVDLS"
            complement(161..194)
            /note="REP01, repeated element"
    repeat_region
        gene
            1253..1336
                /gene="RCRNA24"
            1253..1336
                /gene="RCRNA24"
            1253..1336
                /gene="RCRNA24"
            /product="tRNA-Leu"
            1604..2574
                /gene="mctc"
            1604..2574
                /note="RC0886"
            /codon_start=1
            /transl_table=11
            /product="magnesium transporter"
            /protein_id="AAL03424.1"
            /db_xref="GI:15619992"
            /translation="MPNFNIFKGIILPDHDOFAEIFDQINDLIJEDHNYDAASRLAKL
            HYADLADPILDNLNKKYKTIITLLODEIKKETIVSNAYSKPILIEITLGIKSAEILN
            KLVIEDIIVAVIDLDIDIKDILISNLTKKKEQITVCTTPENTVGRVIRDPDINQO
            GWTVEESLNFIRKVKNDFAIVITDNKLRPIGISMLIKGKONELIKLAKKDFKL
            ADAFTDNLNLSIFEROYALTIIVVYKSGTLVGSVSDINLIYIEEQAGDILISLGG
            HTODTFENVFYTVARPPMLFVNLIITACMTSLIINHNDITAKLVTLATMPDIAVSG
            GNAQTOMVTIVAKLANKDIHYNNVKKVILKEITAVSAGFNGFVLAITIGAGLSFVMLDL
            NLSVIRPAIVIALNFVAGLFGSAPILIHFDIDPATGSGVFLITITDNLGFLTFISSL
            AHIPLV"
            complement(2978..3154)
            /gene="RC0887"
            complement(2978..3154)
            /gene="RC0887"
            /codon_start=1
            /transl_table=11
            /product="unknown"
            /protein_id="AAL03425.1"
            /db_xref="GI:15619993"
            /translation="MFKESNINNHQNIQRDNNISYENNITTFPENQDTSGLCKLGA
            IATVCCNVIVFLFE"
            complement(3316..3482)
            /note="REP06, A and T rich repeated element"
            3823..4275
                /gene="RC0888"
            3823..4275
                /gene="RC0888"
            /codon_start=1
            /transl_table=11
            /product="similarity to (p)pgcP 3-pyrophosphohydrolase"
            /protein_id="AAL03426.1"
            /db_xref="GI:15619994"
            /translation="MROSGNPIYSHPIEIKIMLAEFVAEAPKLYNIMIQVQALLH

```

```

        gene
            complement(4315..4656)
            /gene="RC0889"
            complement(4315..4656)
            /gene="RC0889"
            /codon_start=1
            /transl_table=11
            /product="similarity to proline/betaine transporter"
            /protein_id="AAL03427.1"
            /db_xref="GI:15619995"
            /translation="MVALMSPILNNSNTIELLFOPLFVFAFPIPPAGAVFYARF
            PVLRKPTGSPFIALSRALMFAVSSFRITLIDYNNHGLLELLIPILIGYFVGNHF
            IKLEEAESYV"
            complement(4978..5490)
            /gene="RC0890"
            complement(4978..5490)
            /gene="RC0890"
            /codon_start=1
            /transl_table=11
            /product="similarity to proline/betaine transporter"
            /protein_id="AAL03428.1"
            /db_xref="GI:15619996"
            /translation="MAVILNELFPPEADAPASLYSAIAYCATVYFPGALIFGMIG
            DTIGKATVITITFLIMSLCYLIMANYKNLGNMYGPGSCVGYRSGDVLVRRSSR
            CRALSNRDNKNSSTILGGFACCFNFCNRRVALAVGLTVISFGDLITYCILVMSYC
            YSWCCCKDKS"
            6025..6096
                /gene="RCRNA25"
            6025..6096
                /gene="RCRNA25"
            /product="tRNA-Gly"
            6122..6146
                /note="REP02, RS3-like repeat"
            complement(6239..6397)
            /gene="RC0891"
            complement(6239..6397)
            /gene="RC0891"
            /codon_start=1
            /transl_table=11
            /product="unknown"
            /protein_id="AAL03429.1"
            /db_xref="GI:15619997"
            /translation="MAIFMTVITTSINSELDIIRKAIENYIEKADLLIALSCIEIKR
            KSNFTIRRY"
            6506..6647
                /note="REP05, Rickettsia palindromic element 1"
            6605..6844
                /gene="RC0892"
            6605..6844
                /gene="RC0892"
            /codon_start=1
            /transl_table=11
            /product="unknown"
            /protein_id="AAL03430.1"
            /db_xref="GI:15619998"
            /translation="MMYITANISKVGSQISGEPAERKLIHQRRSLKPDVPKLEVSKYV
            KLLLRGYTRLQDLQNGVTCQOIRKILRIQIITTL"
            6796..7221
                /gene="RC0893"
            6796..7221
                /gene="RC0893"
            /codon_start=1
            /transl_table=11
            /product="unknown"
            /protein_id="AAL03431.1"
            /db_xref="GI:15619999"
            /translation="MSANTQDNQANNNDTIEIOETEVYPAETNSIQSLTSLIPWLI
            FAVFFLLRPDETRKERERKLVSEVKGEVLTSGIYGIYTKVSENDNNIETIETAK
            DVRIYLVKSAIIDIITSRTKEVAVKKNKKVSGAKAS"
            7205..8761
                /gene="secD"
            /note="RC0894"

```

CDS	7205..8761 /gene="secD" /transl_table=11 /product="protein-export membrane protein secD" /protein_id="AAL03432.1" /db_xref="GI:15620000" /translation="MQKLPRKWKIFLSIICTFVAVICALPNFMQVNSFLPHSDVNLGL DLRGALLLDVDDTYLDSNMENLADTLRNFREDKIGYKNLLVRQNSIQLEVRSP KLPEKKIINKIDPEIYAENVENKILSYSSRLNLLNKVVDQSIIVIMRVDSGT SEPTLQGGQKHILLQVGENPSYLNILGKTAKLTFHLVDENATEAVKGVHPVG KMLVQGNASHGEYVVVKKVVLGGDPLTTASAPDQNSOAVAFSNNLGSKIFGE ITKNICKRLAIVLDNKLSPATINGAIMSGSGIITGNFTVESNELALLIRAGSLPA PLKIIERSIPGNLUGAUSIESGKAGLGIYAVCFWWSYGVGLFANIALSLALLY
Query Match	6.0%; Score 112.6; DB 1; Length 10899;
Best Local Similarity	51.7%; Pred. No. 9.4e-15;
Matches 256; Conservative	0; Mismatches 239; Indels 0; Gaps 0;
Qy	284 GGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTATTTGTTGATGAGCGGTATATTAT 343
Db	8270 GGC AAAAGGCGAGGACTTATAGGATTTTATAGCAGTGTGTATATTATGTTGGTCTTTAC 8329
Qy	344 AGATTGGAGCGCTCATCGCTCGGGAGCTGTCTTCTGAATCTTTTCTTATCTGGCA 403
Db	8330 GGTGTGCTGTTTATTTGCCAATATAGCTTAAGCCTTGCATTTATATATTTAGCT 8389
Qy	404 GCTCTACAGTATTTGGATGCGCCACTCACCTTGTGAGGACTCGCTGGGATTTGTTCTGCT 463
Db	8390 TTACTTTTCACTTTTCCAGCTACTTTTAACTCTCGCGGAATTCGAGGAATAACTCACT 8449
Qy	464 ATGGGATGGCGTAGATGCAAAATGTTCTTGATTCGAAAGAAATCCGAGAGAAATTTTA 523
Db	8450 ATGGGTATGGCTGTGTTGCTAAATGTTATGTTTACGAAAGAAATTAAGAGGAACTGCAT 8509
Qy	524 TTGTTCTAAAGTCTTAAATAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCAATT 583
Db	8510 AAAGAGTGTCTACTTTTATGCTATTAGAACCGGTTTGAATTCGCGTTTGGCCACTATC 8569
Qy	584 TTTGATTTCAACTTGACTACAGTATTTGGCTTCCAGCACTTTCTTTCTTCTAGATACAGG 643
Db	8570 CTTGATGCTAACCTAACTACTTTAATCTGTTCTTTTACTTTATATATTGGAGTAGT 8629
Qy	644 CCTATTAAAGGTTTGTGTTGACATGTTATTTAGGAATTTTCTCTCAATGTTTACGGCT 703
Db	8630 GCAATAAAGGTTTGGCGTTGCATTACAGTACGGATTATATCTTCAATGTTTTCGCA 8689
Qy	704 CTTTTCATGACTAAATTTTCTTCTCATGCTTGGATGATAGACCCACATACACAGTTG 763
Db	8690 ATTATCATCTAAATGTTAATAGATATTTGGGTACAATATTTTAAACCTTAAAAAATTA 8749
Qy	764 CATATGATGAATAAG 778
Db	8750 GGGCTTGTTGAAG 8764
RESULT 23	
AE001166	13990 bp DNA linear BCT 15-DEC-1997
LOCUS	AE001166
DEFINITION	Borrelia burgdorferi (section 52 of 70) of the complete genome.
ACCESSION	AE001166 AE000783
VERSION	AE001166.1 GI:2688571
KEYWORDS	
SOURCE	Borrelia burgdorferi.
ORGANISM	Borrelia burgdorferi
REFERENCE	Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.
AUTHORS	1 (bases 1 to 13990) Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Winn, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J., Roberts, K., Bateman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C.

TITLE	Venter, J.C. et al.
JOURNAL	Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi
MEDLINE	Nature 390 (6660), 580-586 (1997)
PUBMED	98065943
AUTHORS	2 (bases 1 to 13990) Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Winn, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J.D., Weidman, J., Utterback, T., Watthey, L., McDonald, L., Artach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C.
TITLE	Direct Submission
JOURNAL	Submitted (12-DEC-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers
source	1..13990 /organism="Borrelia burgdorferi" /db_xref="taxon:139" complement(70..909) /gene="BB0643" complement(70..909) /notes="BB0643" /notes="similar to GB:L43967 PID:1046160 SP:Q49435 percent identity: 38.94; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAC67000.1" /db_xref="GI:2688581" translation="MANKINWFGHKRALDLIKNNLQXANIVLEILDARAPSSKNP LTKITKNQAKILLKHSQVAQINEIILKMKYFENLGNVTIISNYYKGGKQIIDI KKLAIYKKIKYKKEIKVLIGVNNVKGSIINLLSGKSAKVANKPYTKNIOIVKI NESINLFDMPGILMHNLDQSIAKLAIDLMIKNEIVDNTDLALYLEIMDQNNKIL LKKEIYVHKNSLDILQNFARKARKLIGKNELNEKASIKILKEFREGKFGKILDKNY NAT"
gene	1064..1762 /genes="BB0644"
CDS	1064..1762 /genes="BB0644" /notes="similar to GP:1573101 percent identity: 36.77; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAC66999.1" /db_xref="GI:2688580" translation="MLIIIVKIKGLIVSQAENEPHLHSSFIMSKWLAALKIGAGIGI RANGVNDISQIKLEVDLPIIGIKNKNYNCDFITPTMKEDISLCEGVDIIADATP RNRPDVLLDDFFENIKKYPKQCLMADISSDEAINADKGFDFITGLTYGYTKNT GLNIADNDFNLTLLNSLNKSTLIVEKIDTELKAKQCFEMGVDLVVVGATITPAE ITKKFVEKINQIKK"
gene	1808..3352 /genes="BB0645"
CDS	1808..3352 /genes="BB0645" /notes="similar to GB:X80415 PID:1072418 PID:515384 percent identity: 46.03; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="PTS system, glucose-specific IIBC component (ptsG)" /protein_id="AAC66998.1" /db_xref="GI:2688579" translation="MLKGEQAQKFGSRFWLPATILPAAGLLIGIGSLSNPVTPTY SPNIIFLOSQFKIMASGSIIFSNLAPIFISIGIAGLAKSDKGTSGIAAFIYGLVNN ATIGVLIDVSGRAESFSFGVGLTKLETGFGVGVVGLTYLHLSRNFKVDLPK VLGFFSGSRFPVIIVSFSIFLAVIMFLVMPFVQSGINKVGLVDSTGYTGLIYGI



```

LMLAGPGL:HLIYLPPEWTTGGSVIIDGKLEIGTONTFPAELAAGDPRPEIGRS
FMSCGRITWFGFPGALALYTTAKREERTKYGMLSSALNLSFLGTREPLESLF
VAPILTYVATFEDGFAPMLAHIIQITIGTFSSGFVDFTLFGILQNSSTMLNVPYI
GIWVFLYVFTFIFLNFDPFKTGPRTQDINSESSPSSEEEYVARKVLIIGAGA
SNVELDCCATRLRTVRDVLKVSSEKILKTGSKGVIIGKNGQVYVYGVSVLKNEI
EELLEA"
/gene="B3382..4365"
/complement (3382..4365)
/gene="BB0646"
/complement (3382..4365)
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=1
/product="B. burgdorferi predicted coding region BB0646"
/protein_id="AAC67003.1"
/db_xref="GI:2688584"
/translacion="MNINKIIFILFLLILVSPRIKFNKESKLLIPNIEIDNY
LKBELOPNESTTKKEIIMYKKAQKTNSVYVHIGFQSKNEIYVPNNIATAINA
NIFTRKKGHNKNAFRGITTQDNRDIDEAINTGKIDGTVLIGTNGSTAIW
ALANYPNEINSAVLISPNIFPYDKRTNIVYPMGRQIAYLITGYNKFEYKKEH
PTIKSHSRVOHVDAIIMAGLVTLNNSVNFBIKPLIITHPNHTVDPKINFEI
KNYGEKCIPIILLNSHALPIGNQSYSAQNTSYFTKYVPDFINKNK"
/complement (4362..4892)
/gene="BE0647"
/complement (4362..4892)
/notes="similar to GP:1667516 percent identity: 25.19;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="ferric uptake regulation protein (fur)"
/protein_id="AAC66997.1"
/db_xref="GI:2688578"
/translacion="MNDNIIDVHSALEKVGITNDPVLLKNLSELGMAKSHNRNII
VIANPKYETAKEVYVKIKELPSISKATVYTNLTKERNIKDKITDOKETFEY
LSASTIAHRCNKCNOVHPIQDIDIKDLKDLGEMWETKSEIITISGHCNNCKYKD
THNNNVPIENKEITL"
/complement (4992..6680)
/gene="BH0648"
/complement (4992..6680)
/notes="similar to GP:887563 percent identity: 30.22;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="serine/threonine kinase, putative"
/protein_id="AAC66996.1"
/db_xref="GI:2688577"
/translacion="WFRKLKKLTYIIIIISYPSEDNLLIISNIKYLIHQQLAKY
IHWTFPIYFEILREHELMKMLPERKNTDLYMPQTSYSGPHEWMLDEIHDLYW
ALKNPFKSGYKDI.FONTPIFYIYNIIEKPKKTYTELARKLNFYTBGIROSKNNKYL
IFKNNQOULYEVOKIDSPKSNVETLIFYEIKETVNOGLKNFLYLKALLENHDSI
KIONLSSCLUTLELLEIKPNSLKEOPIINPONKRLDOIKESLREPLKHODE
ITKNSBSIYPRKNLEYMNEGFTLSHDVYNIKPNGLKINKIKPKDKVEPEPNTSRYEK
VSSKEILKEASIESFSNSKEKILGKQVLAFNSAKSITIDFIDETISFISIKI
KPNPKSIDLKKATLKKCMPLYLLEYSALIEFPIITGKNTITAKYSDLTGYELIIT
KNPKGYKINDVEFLISKDNKSNFISPLNVEKGIHTINNYIKESKSKWMLING
GSYNTKIDGVINYSQTLNMLPLANNPNKIKLNSKIKNLIIFYNIKKEYENK"
/6875..8312
/gene="B30649"
/6875..8312
/gene="BB0649"
/notes="similar to SP:P27575 GB:X54059 GB:X65139 PID:39365
PID:39374 percent identity: 99.45; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="heat shock protein (groEL)"
/protein_id="AAC66995.1"
/db_xref="GI:2688576"
/translacion="MAKDIYFNEDARKSLISGVKLSNAVKVILGPKRNVLLDKKKG
SPVTYKGVSVAREIELNPFENKGAQLAEVAIKTNIDVAGDGTITATVLAVALAABG

```

```

LKNVSGINPIGKIDHAYNLAAERIKROSARKITTEKEBIAQVASIANDSYIEK
IAEMDKXGKGVITUESKTEPDTISVYEGNOPRGVLSPESTNKEMSVNPDPAF
ILIIYKKSISKEILLPVLEKVLGNTKPLLIITAEDESGALALVNSRGALKVCAIK
SPGFEDRKAMLEDDAVLITGLVLSSEGLTELEVEIQOLGAKTIRYDKDVTIINT
GNKEQIKRSEILKQIEDSTSEYDKELQOERLAVGVGVANVAVTEVELKKEKH
RVEDALSATRAVAEEGVVPGGSLIEVAMYLDTIDTSKLSYEKQGEIYRSLSEEP
MROIISNAGFESGIYHOIKTEKGLAGDASFKVMNIESGIIIDPAKVTBSALONAA
SIAGLLITTECAITDIEKKNKTSGGGVPMDQGM"
/gene="B336..8874"
/gene="BB0650"
/6536..8874
/gene="BB0650"
/complement (336..8874)
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=1
/product="B. burgdorferi predicted coding region BB0650"
/protein_id="AAC67002.1"
/db_xref="GI:2688583"
/translacion="MFLPVNILLFPGIILSDEPFVFCIGDSCRAIVDROLLRENKKG
SAKEIFELGIRYSASAKLYRNDDELINLIEYKINSNTTKDKDRLFGVYVPD
DINKIKVTV"
/8867..9184
/gene="BB0651"
/8867..9184
/gene="BB0651"
/notes="similar to GB:142023 SP:P4592 PID:1003381
PID:1222164 PID:1204499 percent identity: 33.33;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAC66994.1"
/db_xref="GI:2688575"
/translacion="MYSMGGFVFLLOESGNSFLRSLLVFPVIAIEMFLVISPOK
BEKNKEMIKNKKKDKVILIGIGVVKGLDVTIILESPNNEAVPIKNSIDKVL
EKK"
/9270..11030
/gene="BB0652"

```

Query Match 5.9%; Score 110.2; DB 1; Length 13990;  
Best Local Similarity 51.4%; Pred. No. 3.1e-14;  
Matches 281; Conservative 0; Mismatches 263; Indels 3; Gaps 1;

```

QY 183 AACTCGCTCAGATTAAATCGAGCCGATGCTTTGTCGCCGAGTTCGAGGAG 242
Db 10408 ATCTTGCTTAGAGTTAAACGACGCTTTTCAGTATATTAATGATGATTGA 10467
QY 243 AGACGATCTCTCGATCTGGGAAAAACAATGTCACAAAGCATTATCTGACATGCT 302
Db 10468 GAATTAATAGGCTTACTCTTGAGCTAGACATATTGATCTTGACATTAAGCTTCGCC 10527
QY 303 GTGGCTTGGCAATGCTATATGTTTATGAGCGATATATTAAGATTGAGGCGTCATCG 362
Db 10528 TTGCTCTTGTAGTTTTTTTGTATATAGTTATATGTTATATGTTAGAGTGAGTGTGAG 10587
QY 363 CT--TGGGAGCTGTTCTTCTGAATCTTTGCTATCTGGGAGAGCTCTACATATTGG 419
Db 10588 CTGATTTTCACTGTATATTAATGATATTTTAAATTTTAAAGGATATGTGGCCCTTTA 10647
QY 420 ATGGCGCACTCACCTGTGAGACGTCGCTGGAGATGTTCTTGTATAGGGAGTGGCCGTAG 479
Db 10648 ATTTTACTTTTAACCTTAACAGACATTCGAGGCTTATATTTGAACAAAGGATAGCCGTTG 10707
QY 480 ATGCAATGTTCTTGTATGGAAGAATCGAGAGCAATTTTATGTCTCAAAGTCTTA 539
Db 10708 ACAATTAATAGTATATTAAGAGCAATTAAGAGAAATTAAGAGAAAGCAAGATTG 10767
QY 540 AAAAATCTGTAAGAAAAAGATATACCAAGCTTTTGGAGCATTTTGTATCTTAACCTGA 599
Db 10768 AAAATGCTTTTGAAGCTGTTTAAAGGCAATTTTATCAATTAATGATGACAAATTA 10827
QY 600 CTACAGTATGGGCTCAGACCTCTTTCTTCTCAGATAACAGGCGCTATTAAGGGGTTTG 659

```



```

Db 10828 CAACATTTATCGCGTGCTTTTAACTCTCTTGGACAGGATTATCAAGTTTGG 10887
Qy 660 CTTTGACATGATTTTAGGAATTTCTCTTCAATGTTTTCAGCGCTCTTTTCATGACTAAAT 719
Db 10888 CTTGGCTCTCTCTGTCGTAATTTGGCATCCCTTTTATAGTAGTTGATTTTTCAGAT 10947
Qy 720 TTTTCTT 726
Db 10948 TTATTTT 10954

RESULT 24
AE009509/c
LOCUS
DEFINITION Brucella melitensis strain 16M chromosome I, section 66 of 195 of
the complete sequence.
ACCESSION AE009509 AE008917
VERSION AE009509.1 GI:17982603
KEYWORDS
SOURCE Brucella melitensis.
ORGANISM Brucella melitensis
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Brucellaceae; Brucella.
1 (bases 1 to 11540)
DelVecchio,V.G., Kapatal,V., Redkar,R.J., Patra,G., Mujer,C.,
Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A., Lykidis,A.,
Reznik,G., Jablonski,L., Larsen,N., D'Souza,M., Bernal,A.,
Mazur,M., Goltzman,E., Selkov,E., Elzer,P.H., Hagius,S.,
O'Callaghan,D., Letesson,J.-J., Haselkorn,R., Kyripides,N. and
Overbeek,R.
The genome sequence of the facultative intracellular pathogen
Brucella melitensis
Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)
11756688
2 (bases 1 to 11540)
DelVecchio,V.G., Redkar,R.J., Patra,G. and Mujer,C.
Direct Submission
Submitted (13-NOV-2001) Institute of Molecular Biology and
Medicine, University of Scranton, Scranton, PA 18510, USA
3 (bases 1 to 11540)
Elzer,P.H. and Hagius,S.
Direct Submission
Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag
Center, 111 Dalrymple Building, Baton Rouge, LA 70803, USA
4 (bases 1 to 11540)
Kapatal,V., Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A.,
Lykidis,A., Reznik,G., Jablonski,L., Larsen,N., D'Souza,M.,
Bernal,A., Mazur,M., Goltzman,E., Selkov,E., Haselkorn,R.,
Kyripides,N. and Overbeek,R.
Direct Submission
Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell
Park Drive, IL 60612, USA
5 (bases 1 to 11540)
Letesson,J.-J.
Direct Submission
Submitted (13-NOV-2001) Unite de Recherche en Biologie Moleculaire,
Laboratoire d'Immunologie et de Microbiologie, Universite of Namur,
61 rue de Bruxelles, Namur 5000, Belgium
6 (bases 1 to 11540)
O'Callaghan,D.
Direct Submission
Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue
Kennedy, Nimes 30900, France
Location/Qualifiers
1..11540
/organism="Brucella melitensis"
/strain="16M"
/db_xref="taxon:29459"
/chromosome="I"
79..975
/gene="BMEI0674"
79..975

source
FEATURES
gene
CDS

```

```

/gene="BMEI0674"
/codon_start=1
/transl_table=11
/product="SULFATE TRANSPORT SYSTEM PERMEASE PROTEIN CYST"
/protein_id="AAL51855.1"
/db_xref="GI:17982604"
/translation="MARKSRRGVMLPSCFWKNSVLSLMSKRNLSVLPFGFLTGLCTLV
YLTVI VALP LLLAMILKTASIGWDFWNIVTSERALATYRITVSAVAATILNGFLGL
LAWLTRYDFPFGKRLIDAAVDLPALTAIAGLALVTLFQNGWFGVLPFLGKVA
APLGIAMIAPFTSIPFVTVQPVLEDMADVEARS LGATSMQTTHITWPTIFPA
FLAGASLSPARSIGFGSIVFISGNLPFQTEVTALLVIRLDEYDYPAAAAALAFVMLV
MAFFMLAVTNLIQARRLYAER"
962..1792
/gene="BMEI0675"
962..1792
/gene="BMEI0675"
/codon_start=1
/transl_table=11
/product="SULFATE TRANSPORT SYSTEM PERMEASE PROTEIN CYSW"
/protein_id="AAL51856.1"
/db_xref="GI:17982605"
/translation="MQNDAALQARNTPVRNLIWLAAGLSVLCIGAPLAVIFSAYF
SKGIHVPFSEILKPDTLHAYMLTVLTAIVVVPINMAGVCVAMLVTKRPPGRLLIT
FVPIPSVSPIVAGVTYFLYGSQGLLPGLHDKIMETVPAIFLVSLFVTSFPFVA
RELIPLMEVQGSDEEAAALTGASAMQTFYVTLNPKWALLYGAVLCNARVMGEFGA
TSVSGAIRGKNTLPLQVELLFNDYNNVAGAFAPAAASTALIALTVLTKVFLERKQHS"
complement(1838..2083)
/gene="BMEI0676"
complement(1838..2083)
/gene="BMEI0676"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL51857.1"
/db_xref="GI:17982606"
/translation="MVASRNDEYMSFAKAEALSHVWGSGFVDLGHAGHINVASGFGHW
PDGAILASSLHREPAPVNPPLGGLPAPRPLPGWAF"
complement(2129..2443)
/gene="BMEI0677"
complement(2129..2443)
/gene="BMEI0677"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL51858.1"
/db_xref="GI:17982607"
/translation="MPATLLIPGYKGEACHQORWLHDDPSALLVQDDWHYPLVSD
WMHMLEATLAENFGAVLVAHSLGCVLVAHLASRPAAAHVAGALLVAPADAETMARRDS
RF"
complement(2670..3641)
/gene="BMEI0678"
complement(2670..3641)
/gene="BMEI0678"
/codon_start=1
/transl_table=11
/product="LOW PH-INDUCED PROTEIN A"
/protein_id="AAL51859.1"
/db_xref="GI:17982608"
/translation="MKANDYINWPVIGICAVGSANWLLYRELRSISLDDVLSLVAIRM
HHWLVTASALLAYSLSLADYRLALHLRFKISLWLFALCSFTYALSHNIGASVSG
AVVRYAYSQSGMPGSKIAVLIAFCSTFTVLGVIITSSVLLLEPHVLMRPNESLTPA
SVYIALLIMLFFVLLYIFGSMRLPLQIGFSRLEYPRLPVLAQOLVAPLELIGA
IIVFALPEAGNPGFLIILGIFLYSFAALISHAPGGIGLVLELPLTGLPMDQADVLA
ALIVFRLLYLLIIFPAISLVVVLVPEKRSQFVLRWNEKHEQPGKRSR"
complement(3641..5500)
/gene="BMEI0679"
complement(3641..5500)
/gene="BMEI0679"
/codon_start=1
/transl_table=11
/product="POTASSIUM/PROTON ANTI PORTER ROSB"
/protein_id="AAL51860.1"

```

```

/db xref="GI:17982609"
/translation="MPHDPILATITVIGLCIAPIRGATIRUKISPIVGLAGVIAG
PRTGPGVADDDLILOLAETGVLILMPGVGLHSRLKSLVKAIVPAGALAOJIAAAAL
GTGLGLADNDVQAGVFGSLASTASTVILRMQGRLLIDBERGIALGMLIVEDLA
MMVALILPAISVIGRAGQMAADTIVMLGIGGILLTVAKVLIIVMLVGR
KJIPMLNVIAOTGSRRELPRGLVALGVAFGAALFVSPALGAFEPGMSSEL
SHRAEESI.3IRDAFAVLFPISVGMFLDMSLLKDLPLGLTALILLIKSVAAPFIV
RAFRPVSTALTIASLAOIGERSFLLAGVSNILPPAGDILLAGALISFIPNV
MFLVAREMRPMLEKRGATNIDESANTEHNTAEPPPLHTLITGTTITVGGQIGR
LAAIGKEVPELVIEDSPKICISLKEKIEITVAGASNTLIEAEAPVARNVIAI
PNAFEGRVIALGRANSGLHIVRACSOQAQVLEADNSVTLIGEEBIAQMAVAI
GADPERSENVHAPPIEVPAGKEMPEKAEDPATDIIVENGGDMKREBVGIG"
/complete(5578..8115)
/gene="BMEI0680"
/complete(5578..8115)
/gene="BMEI0680"
/codon_start=1
/transl_table=1
/product="PROTEIN TRANSLOCASE SUBUNIT SEC2 / PROTEIN
TRANSLOCASE SUBUNIT SEC2"
/protein_id="AAL51861.1"
/db xref="GI:17982610"
/translation="MRTSKVNVAVLSVLVIGIILAPNFTQKQDALPMLPKRQV
TGGLDRGGSYLVLEVDAAALKKORLTLLDARSKLADRIQPSIRVGDVAVVTI
PDADQAKAETALRTLIQVNTSGPGTAINDIDVTENNUNQILTLTEAGFNRLDSA
LOQSLTIRORVDQVYAEPSIORVSDRIYVQLPGLOPQALROLIGSTAKSPHMV
ADNPNDPEPGVTIMPDSKNPAIKYPIEDQVLSGRLTDAGFDPRTNPIVSFR
FSLGARGOFADITTKVNRFPAILVDGKLSAVIEEPTGSGGQISGFTVDIVL
SALLRAGALPAPLTVIEERTVDPGLGDGAIKQMLMGIGFLVAVFILLVFWPVI
ANVAILLHTLTPSALSIGATLPGIAGIIGIGIAGVAILINERIEETREKIG
AMALDKGHSAPATVDANVTLLTATLITLIFLFGPGVGFATVMMGLGIAISFTLV
LVNIMAMVFRKRLKYVLIETEPKLVPEHTNRPNNAPRIGIVSIVISIIITFP
KPLNTGIGITFGKGOEISTQPADLQOLAKGALNGEVALQOTGSPQOUIRQK
OEGEBEAOJPAINKREAVTELDQVAKIEETEVGSELSAQMLAIVSAVWL
FYIMRPEFEPFALGALATLIDTQKIIIGFPLAQDLPNLAALTLITIGYVNDVY
VYDRMENRLYSKTLREIIMDSINOVLRVCTTFLCPMPAIMGSAVHNFA
VPMFLGVIATSSSIFAPILLDLDWMQRRHNAKHTAEGALAK"
/complete(8224..8640)
/gene="BMEI0681"
/complete(8224..8640)
/gene="BMEI0681"
/codon_start=1
/transl_table=1
/product="Hypothetical Protein"
/protein_id="AAL51862.1"
/db xref="GI:17982611"
/translation="MTGRKKTEDDRQHGMLFVTCVLSFTAEFLFTVDFSSQAGDALN
OPRGIAROEGSAAPAROOSTPAPQIPRAIMEAIAAKIKTVMIGDNAAALLTQTA
ELHVSNGRBAEAAQIPQPLALRGHFSARAPPENNA"
/complete(8962..11457)
/gene="BMEI0682"
/complete(8962..11457)
/gene="BMEI0682"
/codon_start=1
/transl_table=1
/product="mechanosensitive ion channel"
/protein_id="AAL51863.1"
/db xref="GI:17982612"
/translation="MAQNDKQONPASQAPAEKPNQQAASADQRPPEVSPVSI
VOOORPEDEBLQKTKINOLSKSASDEALNLIKQDALSKLLETAVRPRILT
EINRLEQLGAPSSDQAPAPAIYGERARLLAEKAINNTIGETETSTAVNRMSLL
IGDMRDLFTKTLSDQVNDSTLGSSEVYASADOMISLRIYSSMRFPVTFKLESL
AAAFALAAALVLOFGARFAGAYRDPVESPSYSTRISAVSWSTVISAQVAVL
ATTYFLINFLVLTLDIASLFOSLFIYGLVFFHRLAVACISDMQOMLVQVAPP
GHLLAMVTATLSGDSFGVNRILISSPLSTMAKSIATVITGLIALLAVFVP
VERDEKAVAMPAPRIPLIMGLILPILALSGVIGIABPISQOIVVGAFLVMMV
GFLTKRAISEBOAFASRRIGKAKERHHPBATIDQGLAGILINLVALLIGPVL
MQLGFQWAEIKSTFYKLTMTGFIGNFSISLMLGSLVLLPLIGVTLTRMQLNDNSV

```

Query Match 5.9%; Score 109.6; DB 1; Length 11540;  
Best Local Similarity 49.8%; Pred. No. 4,3e-14;  
Matches 277; Conservative 0; Mismatches 279; Indels 0; Gaps 0;

QY	150	TCTCAGGAAATTTTACCCACCGTGAAGTGAAGCAAACTCGCTCAAGATTTAAATCTGAG	209
QY <td>7151</td> <td>TACGCGGAGCTTACACCGTTACAGATAGCGGTGTCTCTCCGGGCTTCGCGGAGGCG</td> <td>7092</td>	7151	TACGCGGAGCTTACACCGTTACAGATAGCGGTGTCTCTCCGGGCTTCGCGGAGGCG	7092
QY <td>210</td> <td>CGATGCTTTTGTTCGCGAGTTCAGTGAAGACGATCTCTTGTATCTTGGGAAA</td> <td>269</td>	210	CGATGCTTTTGTTCGCGAGTTCAGTGAAGACGATCTCTTGTATCTTGGGAAA	269
QY <td>7091</td> <td>CCCTGCGCCCTCGCTCACCGTCATCGAAGAACGACCGCTGCGCCGAGATCTTGGGCG</td> <td>7032</td>	7091	CCCTGCGCCCTCGCTCACCGTCATCGAAGAACGACCGCTGCGCCGAGATCTTGGGCG	7032
QY <td>270</td> <td>AACATGTACAAAGGCAATATCTGACAGTGTGAGTGTGGCAATGCTTATGTTTGA</td> <td>329</td>	270	AACATGTACAAAGGCAATATCTGACAGTGTGAGTGTGGCAATGCTTATGTTTGA	329
QY <td>7031</td> <td>ACGCCATCAAGATGGGCTGATGACAGATATATCGGCTTCTTCGTCGCGGCTTCA</td> <td>6972</td>	7031	ACGCCATCAAGATGGGCTGATGACAGATATATCGGCTTCTTCGTCGCGGCTTCA	6972
QY <td>330</td> <td>TGAGCGTATATATATAGATTTGAGAGCGTCATCGCTTCGGAAGCTGTTCTTGAATCTT</td> <td>389</td>	330	TGAGCGTATATATATAGATTTGAGAGCGTCATCGCTTCGGAAGCTGTTCTTGAATCTT	389
QY <td>6971</td> <td>TCCGTGCTTCTATGATTCGCGGCTGATTCGGAATTCGCGGCTTCTGCTCATACCG</td> <td>6912</td>	6971	TCCGTGCTTCTATGATTCGCGGCTGATTCGGAATTCGCGGCTTCTGCTCATACCG	6912
QY <td>390</td> <td>TGCTTATCTGGGAGCTCTACAGATTTGAGTGGCCACTCACCTTGTCAAGACTGCTG</td> <td>449</td>	390	TGCTTATCTGGGAGCTCTACAGATTTGAGTGGCCACTCACCTTGTCAAGACTGCTG	449
QY <td>6911</td> <td>TGCTGACCTTCTCGGCGCTAGCGCTGATCGGCGCAACGCTGACGCTGCCGATGCGAG</td> <td>6852</td>	6911	TGCTGACCTTCTCGGCGCTAGCGCTGATCGGCGCAACGCTGACGCTGCCGATGCGAG	6852
QY <td>450</td> <td>GGATGCTTCTGCTATGGGAGTGGCGGTAGATGCAAAATGTTCTTATCGAAAGATCC</td> <td>509</td>	450	GGATGCTTCTGCTATGGGAGTGGCGGTAGATGCAAAATGTTCTTATCGAAAGATCC	509
QY <td>6851</td> <td>GCAATATCTCGGATCGGATCGGCTGATGCGAATTTCTCATCAACGAGCGATCC</td> <td>6792</td>	6851	GCAATATCTCGGATCGGATCGGCTGATGCGAATTTCTCATCAACGAGCGATCC	6792
QY <td>510</td> <td>GAGAGAAATTTTATTTGCTCAAGTCTTAAATAATCTGTAGAAAAAGATATACCAAG</td> <td>569</td>	510	GAGAGAAATTTTATTTGCTCAAGTCTTAAATAATCTGTAGAAAAAGATATACCAAG	569
QY <td>6791</td> <td>GTAAGAAACCCGCAAGGACCTTGCGCGATGCGCGGCGCTCGACAAAGGCTTCCACAGG</td> <td>6732</td>	6791	GTAAGAAACCCGCAAGGACCTTGCGCGATGCGCGGCGCTCGACAAAGGCTTCCACAGG	6732
QY <td>570</td> <td>CTTTGGAGCATTTTGTATTTCTAATCTGACATGATTTGCTTGAACCTTCTTTCT</td> <td>629</td>	570	CTTTGGAGCATTTTGTATTTCTAATCTGACATGATTTGCTTGAACCTTCTTTCT	629
QY <td>6731</td> <td>CTTTGCGACCATCGGAGCGCAATATGACGACGCTGACGCGCAACATCTTCTGTTCC</td> <td>6672</td>	6731	CTTTGCGACCATCGGAGCGCAATATGACGACGCTGACGCGCAACATCTTCTGTTCC	6672
QY <td>630</td> <td>TCTTACATACAGGCGCTATTAAGGCTTGTGACATTTATAGAAATTTCTCT</td> <td>669</td>	630	TCTTACATACAGGCGCTATTAAGGCTTGTGACATTTATAGAAATTTCTCT	669
QY <td>6671</td> <td>TGTTGCGACCGGCGGCTTGGCGCTTGGCGCTCATGATGCTTGGATGCCATTT</td> <td>6612</td>	6671	TGTTGCGACCGGCGGCTTGGCGCTTGGCGCTCATGATGCTTGGATGCCATTT	6612
QY <td>690</td> <td>CAATGTTACGCTCT 705</td> <td></td>	690	CAATGTTACGCTCT 705	
QY <td>6611</td> <td>CGATGTTACCGATGT 6596</td> <td></td>	6611	CGATGTTACCGATGT 6596	

RESULT 25  
MPUM03/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
SOURCE

MPUM03 315079 bp DNA linear BCT 14-MAY-2001  
Mycoplasma pulmonis (strain UAB CTIP) complete genome; segment 3/3.  
AL445565 AL445566  
AL445565.1 GI:14089942

Mycoplasma pulmonis.  
Mycoplasma pulmonis  
Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
1 (bases 1 to 315079)  
Chambaud, I., Heilig, R., Ferris, S., Barbe, V., Samson, D.,  
Gallison, F., Moszer, I., Dydyk, K., Wroblewski, H., Viat, A.,  
Rocha, E.P.C. and Blanchard, A.  
The complete genome sequence of the murine respiratory pathogen  
Mycoplasma pulmonis  
Nucleic Acids Res. 29 (10), 2145-2153 (2001)

Journal  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
SOURCE

Submitted (16-OCT-2000) Blanchard A., INRA, Centre de Recherche de  
Bordeaux, Institut de Biologie Vegetale Molculaire, 71 avenue  
Edouard Belin, BP 81, 33883, Villenave d'Ornon, Cedex, France  
Location/Qualifiers  
1..315079  
/organism="Mycoplasma pulmonis"  
/strain="UAB CTIP"  
/db\_xref="taxon:2107"  
288..2807  
/gene="MPYU\_5280"

```

CDS
288..2807
/standard_name="vsai"
/notes="identified by sequence similarity; similar to
TREMBL:Q50279 Blastp2 P=160 Id=45% ; Pfscan: pos. 20-30
PS00013 | PROKAR LIPOPROTEIN Prokaryotic membrane
lipoprotein lipid attachment site ; Gene Rearrangements
in the vsa Locus of Mycoplasma pulmonis; Xuejun Shen,
Juliann Gumulak, Huilan Yu, C. Todd French, Nianxiang Zou,
and Kevin Dybvig; J. Bacteriol. 2000; 182 2900-2908"
/codon_start=1
/evidence=experimental
/transl_table=4
/product="LIPOPROTEIN VSAI"
/protein_id="CAC13701.1"
/db_xref="GI:14089943"
/translations="MKNSSMKKFLFLGSLTSLFIAGATLIACQKODGAIKKDLSES
KIDVLPFKPSVDETAQWTVRLSEYRTAKSSDSNLTDFAFVKSGLSGSDYELT
FKKGADKIEVITYKSKAVSSTKVLEDTKIGOSDAQYNALDLRVKSLVANSDDLK
GKPSVEVIEKTSVDVKETKQADVDVFTNMGTVITKSSASDAADAKGLKVSV
QFSGDQFKSKDLMLTGLKTSNNQGGNNGSGSKPMTPPNKNGSGDSMENPWTPE
KNGSDSMENPMTPPPPNNQGGNNGSGSKPMTPPNKNGSGDSMENPWTPE
NPOGQMDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQ
POGQMDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQ
QNGQMDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQ
NGQMDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQ
QMDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQ
QMDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQ
MDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQ
DNQNDMGNTNPOGQMDNQGNDMGNTNPOGQMDNQGNDMGNTNPOGQMDNH"
complement(3945..4073)
/gene="MYPU 5290"
complement(3945..4073)
/gene="MYPU 5290"
/standard_name="vsah"
/notes="identified by sequence similarity, similar to
trEMBL:Q50322 blastp2 P=2e-8 score=55; Gene Rearrangements
in the Locus of Mycoplasma pulmonis; Xuejun Shen,
Juliann Gumulak, Huilan Yu, C. Todd French, Nianxiang Zou,
and Kevin Dybvig; J. Bacteriol. 2000; 182 2900-2908"
/codon_start=1
/evidence=experimental
/transl_table=4
/product="LIPOPROTEIN VSAH (FRAGMENT)"
/protein_id="CAC13702.1"
/db_xref="GI:14089944"
/translations="GGSGSGSTNNPMTPTTAPGSGSGKTETPTTADTPAGARR"
4331..6067
/gene="MYPU 5300"
4331..6067
/gene="MYPU 5300"
/standard_name="lipA"
/notes="identified by sequence similarity; possible;
similar to TREMBL:Q50274 (Q50274) Blastp2 P=0 C=38% ;
Pfscan: pos. 14-24 PS00013 | PROKAR LIPOPROTEIN
Prokaryotic membrane lipoprotein lipid attachment site ;
Gene Rearrangements in the vsa Locus of Mycoplasma
pulmonis; Xuejun Shen, Juliann Gumulak, Huilan Yu, C. Todd
French, Nianxiang Zou, and Kevin Dybvig; J. Bacteriol.
2000; 182 2900-2908"
/codon_start=1
/evidence=not experimental
/transl_table=4
/product="LIPOPROTEIN A"
/protein_id="CAC13703.1"
/db_xref="GI:14089945"
/translations="MNNKYFKYSWVLLSTSLAPMTLASCNHNVAKKEDTKQDSS
NLNKNKTSNDPHLKDKNVQDNKDSNKAVSNSQTSQKTNESQTKDQSS
KTNLIITNQNSSNTSKIOENKQSDQNTSAVNSALEKOTKNDENISLVNSKDTN
VILNKDEKVALAKDSKSKSNLNLTPTVENRONKNEVKDDKKALOWOKLNSA
SILESFSYDQTSLSLTFKEMGNPLGLEVVLEKLNDSHEEKEISFKTNGKVNLLTS
SNLTSKGWKIKSFSDKTSFHSPTIETTFDFKTNELKQRIEIKQAIKDYQIKQ
```

```

gene
complement(7286..8110)
/gene="MYPU 5310"
CDS
complement(7286..8110)
/gene="MYPU 5310"
/notes="identified by sequence similarity; putative;
similar to SWISS-PROT:P19776 (CDV_BACSU) Blastp2 P=2e-11
C=30% ; Gene Rearrangements in the vsa Locus of Mycoplasma
pulmonis; Xuejun Shen, Juliann Gumulak, Huilan Yu, C. Todd
French, Nianxiang Zou, and Kevin Dybvig; J. Bacteriol.
2000; 182 2900-2908"
/codon_start=1
/evidence=not experimental
/transl_table=4
/product="INTEGRASE/RECOMBINASE"
/protein_id="CAC13704.1"
/db_xref="GI:14089946"
/translations="MFCFFILTKIVFLSKPKLTGENMIEKYCAPLEKRNLSKKY
VDSRRILSKLDVENNSHRKIMKIIVDSNGAHVORMLLASVKGFLKFKHYKVED
LILFKTKYDVIYRVPVLKRLKLTLPENDSERIKTKILIRFLYOTGIRIGELNT
LILVNKKLYVHGKKNRQILYLEETFTFNRYIPDLRYPMSLKTLRIEIKLIGKEF
SPHSRLRSFATHMMQSGADPKTLMQLGHSSINTTFQYVNSSESNNRKYIKHLNK"
complement(8147..9031)
/gene="MYPU 5320"
complement(8147..9031)
/gene="MYPU 5320"
/standard_name="tsf"
/notes="identified by sequence similarity; probable;
similar to SWISS-PROT:P19216 (EFTS_SPICI) Blastp2 P=3e-67
C=63%
/codon_start=1
/evidence=not experimental
/transl_table=4
/product="ELONGATION FACTOR TS (EF-TS)"
/protein_id="CAC13705.1"
/db_xref="GI:14089947"
/translations="MDASOKAGLIIKKLREITNSGFLDCKKALBETNYDLDAEMLOE
NGKAKAAKSGRIAAEGLVRASVKGSAVIFELNSETDFVARNKEFLDMDNISEALV
ENSFQSLSAENIFMENDLTLEATTATATIGEKISFRRAKFDLLEDQTIQAYTHA
NGRIASLFLVRGNEEVAKNVAMHIAAMNPEYMANEVPQEKIEKKAEPKSPALAG
KPEKIQOISILSGMLNKAELAEFVLLNQPFVMESSLSVEQYLKNNKSEALEMIRYEVGEG
IEKKAVDFASEVAQMKK"
complement(9034..10026)
/gene="MYPU 5330"
complement(9034..10026)
/gene="MYPU 5330"
/standard_name="rpsB"
/notes="identified by sequence similarity; possible;
similar to SWISS-PROT:P21464 (RS2_BACSU) Blastp2 P=1e-63
C=50%
/codon_start=1
/evidence=not experimental
/transl_table=4
/product="30S RIBOSOMAL PROTEIN S2"
/protein_id="CAC13706.1"
/db_xref="GI:14089948"
/translations="MAEKLRSKNEQKQVNTENIYKIIIPNSKFEAGVYFGRHTQOWN
QAOQRTNSYVYVTERWLTGTLTNSKTIISVRRLDLEKMAENNFEGYTKKGIKPEKEL
SKLKNFEGIREMKTQPMFVVDPNVRIAVKEADKGVKIIGILSDSDANPFDVLG
IPANDDSIKSISLILVLADAINVAKGKDLFPAYQDEANILPEDKIDSPQSHFNN
TRPDQSSREHDSWRNSQNTSQNOEQSRPYKRTYKNSDRETLDDKTKTBEVK"
10241..10732
/gene="MYPU 5340"
10241..10732
/gene="MYPU 5340"
/notes="no similarity found"
/codon_start=1
/evidence=not experimental
```

Query Match 5.8%; Score 108.8; DB 1; Length 315079;  
 Best Local Similarity 49.5%; Pred. No. 4.6e-16;  
 Matches 281; Conservative 0; Mismatches 287; Indels 0; Gaps 0;

153 CAGGAAATTATCCACCGTGAAGTGAACAACCTGCTCAGATTATTAATCTGAGAGA 212  
 Db 143811 CTGTAATTATTTACTGCTCTTACCTAACAACCTGCTGAGTAACTTGGTTCAG 143852

QY 213 TGTCTTTTGTCCGAGGTTCTCAGTGAAGAGAGCATCTTCTTGATTTGGGAAAAAC 272

Db 143851 GTGATTAACATTTAGAGGCTTTATCTTCAGATTGTGTGAAGCTTATGAACTCAT 143792

QY 273 AATGTACACAAGCATTTATCTCAGATGCTGCTGGCAATGCTTATGTTTGAATGA 332

Db 143791 CATTTAATGCAATGCAATTTATGAGAAATTAATTTCAATTATGTAATGCTTAATATGA 143732

QY 333 GCGTATATTATAGATTGGAAGCGCATGCTTCGGAGAGCTGTTCTTCTGATCTTTTGC 392

Db 143731 TTTACAACTATGCTGTTCTTGAATAGCTCTCTTGTCCATGAGATTATTAATATCC 143672

QY 393 TTAATCGGAGCTCTTACAGTAATTTGATGCGCACTCACTTGTGAGAGCTCGTGGGA 452

Db 143671 TTTCACTAACTGTTTACTTGATTAATGAGAAATTAATCTCAAGCTCAATAGCGGCTT 143612

QY 453 TGTCTTCTGCTATGAGGATGCGCCGATGAGAAATGTTCTTGATTTGAAAGAAATCCGAG 512

Db 143611 TAGTAATAGCTATAGGATATGTCAGTTGATGCGCAATTAATTAATTTGAAGATTAAC 143552

QY 513 AGGAATTTTATGTCCTGAAGTCTTAATAAATCTGTAGAAAAGATATACCAAGGCTT 572

Db 143551 AGCACTCAAAATATGAGAAATGCTTAATAAAGCTTAATAAATGCTCAATGATTTCTCTC 143492

QY 573 TTGAGCACTTTTATGATCTTAATGCTTACAGTAATGAGCTCAGCACTCTTTTCTTCC 632

Db 143491 TTTCAACAATTTTGGAGCTCAAACTAATCTTAATTTGATGAGAAATCTTTTCTTCT 143432

QY 633 TAGATACAGGCGCTTATTAAGGTTTGTCTTGAACATGATTTTAGAATTTTCTCTCA 692

Db 143431 TTGGTACAAAGCTCATTTAGAGGTTTGAAGATCTCACTTATCTCTGCAATATTAATTT 143372

QY 693 TGTATACGCGCTTTTCAATGACTAATTT 720

Db 143371 TGTATGATGTTAATTAATCACTAATTT 143344

RESULT 26  
 AX084084  
 LOCUS AX084084 1848 bp DNA linear PAT 28-FEB-2001

DEFINITION Sequence 119 from Patent WO0111033.  
 ACCESSION AX084084  
 VERSION AX084084.1 GI:13185634  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Haemophilus influenzae.  
 Haemophilus influenzae.  
 Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 Haemophilus.  
 REFERENCES  
 1 (bases 1 to 1848)  
 Chovan, L.E., Heseler, P.B. and Reich, K.A.  
 Essential bacteria genes and genome scanning in Haemophilus  
 influenzae  
 Patent: WO 0111033-A 119 15-FEB-2001;  
 ABBOTT LABORATORIES (US)  
 FEATURES  
 Location/Qualifiers  
 1..1848  
 /organism="Haemophilus influenzae"  
 /db\_xref="taxon:727"  
 1..>1848  
 /note="H1-0240"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAC33403.1"  
 /db\_xref="GI:13185635"  
 /translation="MLNRPYLNKMLVFIYAIGILYSLPNYGEDPAVQISCTGQE  
 ANTSLVGVQVDVLTNNLPTKSVLENGSILAFNTDQILAKOKLARLNNTTA  
 LNLAPATPMLSMFGANPMKGLDVGVRFLMDMNAATLVKROBODSDLRGELRK  
 EKIOYTAIENHFGTLITLAVSOKAKERIRIOLPLDITDEPDADISNGLSTAA  
 LNEARDIAEONLTLIRKVAELGVAERIORGAEIRIYELIEPDRRAKEIGAT  
 ATLERIYNQNTAALISRNMLPADSEVYDQGHVLPKRAVGGEEHINSSGLD  
 QHSSTPQVSVTLDSGGEIWSQTKTKYIKPMATLVEYKDKDDEKNTLLEKHE  
 VINVAITIQGRFGSNFQITGVDSIAEAKHNLSTLKSGLALPIQIVERTGSLGQON  
 VEOGINASLMGLVAIVAFMLFYKMGVSLPALVINIVLGLVMSILGALSMGI  
 AGIVTLGMSVPAVNLIFERIKELIRNGSIOQALINEGNGAFSTIFDNLTLTLTAI  
 ILVAVGTGIGQFAITLSLGAISMETAITGRALVNALYKQJLKLLI"  
 BASE COUNT 560 a 321 c 368 g 579 t  
 ORIGIN

Query Match 5.8%; Score 108.2; DB 6; Length 1848;  
 Best Local Similarity 51.4%; Pred. No. 1.1e-13;  
 Matches 276; Conservative 0; Mismatches 258; Indels 3; Gaps 1;

QY 173 GAAGTGACAACTGCTCAGATTAAATCTGAGAGATGCTTTTGTCCGAGGT 232

Db 1249 GAAGCATATATCTTTCTACTTATTAATAATCTGAGCATTAATGACCAATTCAAAT 1308

QY 233 CTAGTGAAGAGCATCTCTTCTGATCTTGGGAAAAACAAGTACACAAGGATATC 292

Db 1309 GTTGAAAGACGCAATGCGCCATCATTAAGTGGCGCAAACTGAGCAAGGATTAAT 1368

QY 293 TCACATGCTGTGGCTTGGCAATGCTTATGTTTGAAGCGTATTAATTAATTTGA 352

Db 1369 GCGAGTCTTTGGGATTAATGCTGTTATGCTTTATGTTTATCTCAAAATGTTT 1428

QY 353 GCGCTCATCGC--TTGGGAGCTGTTCTTGAATCTTTTGTCTTATCTGGGAGCTCTA 409

Db 1429 GGTGTGATGGAAGTTTGAACCTGTTATTAATATCTGATTAATCTTGGGATTAATGCT 1468

QY 410 CAGTAATTTGATGCGCACTCACTTGTCAAGATCTGCGTGGAGATGTTCTTGAATGGG 469

Db 1489 ATTTTACCGGCGGACACTTCAATGCCGGTATTCGGGATTCGTTTAACTTTAGGT 1548

QY 470 ATGGCCGATGATGAATGTTCTTGATTTGAAAGAAATCCGAGAGAAATTTTATGTC 529

Db 1549 ATGTCAAGTATGCGAATGATTAATTTTGAAGTATTAAGAAAGAAATTCGTAATGCT 1608

QY 530 CAAAGCTTAATAAATCTGTAGAAAAGATATACCAAGCTTTTGAAGCATTTTGTAT 589

Db 1609 CGTTCAATTTGCAAGCATTAATGAAGTATTAACGCGCATTTACTTCAATTTTGTAT 1668

QY 590 TCTAATCTGATCAAGTATTTGGCTTCAGCACTTTCTTTCTTCTTGAATACAGGCGCTAT 649

Db 1669 GCAAACTTAACCAATCTTAACCGCAATTAATTTCTATACGCGGTAGAACAGGCGCAAT 1728

Qy 650 AAAGGTTTCTTGCATGATTTTAGGAATTTCTCTTCTTCAATGTTTACGGCTTT 706  
 Db 1729 CAAGGTTTTCGATTCAGCTTTCACCTTGTGTGGATTTCTTCTATCTGTATCCCGGATT 1785

RESULT 27  
 AX395145  
 LOCUS AX395145 1848 bp DNA linear PAT 18-MAY-2002  
 DEFINITION Sequence 119 from Patent WO0218601.  
 ACCESSION AX395145  
 VERSION AX395145.1 GI:21066198  
 KEYWORDS  
 SOURCE Haemophilus influenzae.  
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 Haemophilus.

REFERENCE 1  
 AUTHORS Chovan, L.E., Hessler, P.E. and Reich, K.A.  
 TITLE Essential bacteria genes and genome scanning in I Haemophilus inf  
 JOURNAL Patent: WO 0218601-A 119 07-MAR-2002;  
 Abbott Laboratories (US)

FEATURES  
 source Location/Qualifiers  
 1. 1848  
 /organism="Haemophilus influenzae"  
 /db\_xref="taxon:727"  
 1. >1848  
 /note="HI-0240"  
 /codon\_start=1  
 /transl\_table=1  
 /protein\_id="CAD32145.1"  
 /db\_xref="GI:21066198"  
 /translation="MLNRYPLWKNLWIFVAIGILYSLNPYIGEDPAVOISGTRGOE  
 ANTSLVGQVODVLKTNLPKTSIVLENGSILARTNTDDOLAKDKIAERLGNNTYA  
 LNLAPATPAWLSMFGANPMKGLDRCGVRELFMEVDMNATLVKQEOLODSLGELEK  
 EKIOYAKNTEHSGTILTLANVSORAKAERIIRQLPTLIDTEPDADSLNGLSTAA  
 LNEARDLAEONLILKRVAELGVAERIQORGAERIVILPGVQDTARAKELIGAT  
 ATLEFRIVQNTADALSRNMLPADSEVKYDROGHPALFKRVLGGEHINSSGLD  
 QHSPTQVTLSEGEIINSQTKYKPKMATLYVYKDKNGKKBENGKILKEHE  
 VINVATTQGRFSNQITGVDSIAEHNLSFLKSGALIAPIQIVERTIGPSLGAQN  
 VEQGINASLWGLVAVIAPLFPYKMGVIAISFALVINIVLLGLMSILPGATLSMPCI  
 AGIVLTGMSVDANVLIFERIKERNGRSIQQAINESYNGAFTSIFDANLITLITAI  
 ILVAVGTGPIGGFNAITLSLGVASIMFATITGTRALNVALYGGKOLKLLI"

BASE COUNT 560 a 321 c 388 g 579 t

Query Match 5.8%; Score 108.2; DB 6; Length 1848;  
 Best Local Similarity 51.4%; Pred. No. 1.1e-13;  
 Matches 276; Conservative 0; Mismatches 250; Indels 3; Gaps 1;

Qy 173 GAAGTGAGCAACTCGCTCAGATTAAATCTGGAGCGATGCTTTTGTCCCGAGGTT 232  
 Db 1249 GAAGCACATAATCTTTTACCTATTGAAATCTGTGCGATTAAATGCACCAATTCAAATT 1308

Qy 233 CTCAGTGAAGAGACATCTTCTGATCTTGGGAAAAACAATGTACACAGGCATATTC 292  
 Db 1309 GTTGAAGAACGCACAAATTTGGCCCATCATTAGGTGCGCAAAAGTAGACAGCGGATTAAT 1368

Qy 293 TCAGCATCTGTGGCTGCGCAATGCTTATGTTTGTGAGCGGTATATATAGATTGGA 352  
 Db 1369 CGGAGCTTTGGGGATAGTGTCTGTTATGTCCTTTATGCTTTTACTACAAAATGTTT 1428

Qy 353 GCGCGTCATCGC---TTCCGGAGCTGTTCTTCTGAATCTTTTCTTATCTGGGCGACTCTA 409  
 Db 1429 GGTGTGATTGCAAGTTTTCGACTTGTATTAATATCGTATTACTTGTGGGATTATGTCT 1488

Qy 410 CAGTATTTGGATCGCCCACTACCTTGTACGACTCGCTGGGAATGTTCTTGTCTATGGGG 469  
 Db 1489 ATTTTACCGCGCGCACACTTTTCAATCGCGGTATTCGGGGTATCGTTTAACTTTAGGT 1548

Qy 470 ATGCCCGTAGCAAGTCTTCTGATTCGAAAGAACCCGAGAGAAATTTTATGTCT 529  
 Db 1549 ATGTCAGTAGAGCGGAATGTATGTATTTTGAACGATTAAGAAAGAAATTCGTAATCGT 1608

Qy 530 CAAAGTCTTAAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTAT 589  
 Db 1609 CGTTCAATTTCAGCAAGCCATTAAATGAAGGTTATAAGCGCGCATTTACTTCTATTATTGAT 1668

Qy 590 TCTAACTTGACTACAGTATTGGCTTCAGCACTTCTTTTCTTCCTAGATACAGGCGCTATT 649  
 Db 1669 GCAAACCTTAAACCAATCTTAAACCGCAATTTCTATACGCGGTAGGACAGGCGCAATT 1728

Qy 650 AAAGGTTTGTCTTGGACATGATTTTAGGAATTTCTCTTCAATGTTTACGGCTCTT 706  
 Db 1729 CAAGGTTTTCGATTCAGCTTTCACCTTGTGTGGATTTCTATGTTTACCGCGATT 1785

RESULT 28  
 U32710/c  
 LOCUS U32710 13818 bp DNA linear BCT 29-MAY-1998  
 DEFINITION Haemophilus influenzae Rd section 25 of 163 of the complete genome.  
 ACCESSION U32710 L42023  
 VERSION U32710.1 GI:1573200  
 KEYWORDS  
 SOURCE Haemophilus influenzae Rd.  
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 Haemophilus.

REFERENCE 1 (bases 1 to 13818)  
 AUTHORS Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A.,  
 Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J., Dougherty, B.A.,  
 Merrick, J.M., McKenney, K., Sutton, G.G., FitzHugh, W., Fields, C.A.,  
 Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.I., Glodek, A.,  
 Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E.,  
 Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, D.T., Saudek, D.M.,  
 Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L.,  
 Geoghegan, N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M.,  
 Smith, H.O. and Venter, J.C.  
 TITLE Whole-genome random sequencing and assembly of Haemophilus  
 influenzae Rd  
 JOURNAL Science 269 (5223), 496-512 (1995)  
 MEDLINE 95350630  
 PUBMED 7542800

REFERENCE 2 (bases 1 to 13818)  
 AUTHORS Tatusov, R.L., Muehlegan, A.R., Bork, P., Brown, N.P., Hayes, W.S.,  
 Borodovsky, M., Rudd, K.E. and Koonin, E.V.  
 TITLE Metabolism and evolution of Haemophilus influenzae deduced from a  
 whole-genome comparison with Escherichia coli  
 JOURNAL Curr. Biol. 6 (3), 279-291 (1996)  
 MEDLINE 96398784  
 PUBMED 8805245

REFERENCE 3 (bases 1 to 13818)  
 AUTHORS White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-JUL-1995) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA

REFERENCE 4 (bases 1 to 13818)  
 AUTHORS White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-SEP-1997) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA

REMARK The H. influenzae sequence has been updated by R. Fleischmann. New  
 database matches have been assigned, product names have been  
 improved, and a number of frame shifts have been corrected. We  
 gratefully acknowledge the work of Tatusov et. al. We have  
 incorporated their annotation into the /notes fields of the  
 corresponding H. influenzae genes

REFERENCE 5 (bases 1 to 13818)  
 AUTHORS White, O., Clayton, R.A., Kerlavage, A.R., Fleischmann, R.D.,  
 Peterson, J., Hickey, E., Dodson, R. and Gwinn, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-MAY-1998) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA

REMARK The whole genome was shifted by 588 nucleotides for a new start  
 On Sep 30, 1996 this sequence version replaced gi:1220911.  
 Location/Qualifiers

source 1.13818  
/organism="Haemophilus influenzae Rd"  
/db\_xref="taxon:71421"  
complement(134.343)  
gene /gene="HI0235"  
complement(134.343)  
CDS /gene="HI0235"  
/note="similar to GB:129458 SP:P36675 PID:473421 percent identity: 73.81; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAC21905.1"  
/db\_xref="GI:1573201"  
/translation="MRKQKSAVENETVYVLTFRGVKIDNAVALLQDKLFRQREKKR KGGSYQRAKHPGKIFKPDYKF"  
complement(411.761)  
gene /gene="HI0236"  
complement(411.761)  
CDS /gene="HI0236"  
/note="similar to PID:1805555 percent identity: 45.74; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="arsenate reductase, putative"  
/protein\_id="AAC21906.1"  
/db\_xref="GI:1573202"  
/translation="MSVILYHNPHSKSRRETLLENKGIQPIIEYLQKQSVNELQ STAKKIGIDVRQMRKIDELYSIMLDNLQAELEKAISSHSALIERPIVINDK AKGRPEPEVLEIL"  
840.1868  
gene /gene="HI0237"  
/note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:U00096 SP:P77406 PID:1621505 PID:1788838 percent identity: 55.75; identified by sequence similarity; putative"  
/pseudo  
complement(2003.2980)  
gene /gene="HI0239"  
complement(2003.2980)  
CDS /gene="HI0239"  
/note="similar to SP:P19674 GB:X56175 PID:581231 GB:U00096 PID:1773093 percent identity: 48.00; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="protein-export membrane protein (secf)"  
/protein\_id="AAC21907.1"  
/db\_xref="GI:1573204"  
/translation="MMKLFYKDXDGHFIREINGIKLPPLTEPMKVKRYKYLISALLM VISLFFITKGMNGDFTGVVDFTFSQSANLEQSRSLHNGESSPIVQTTGSVO DNMIRLPASNNDSITIGHVSXKLONVDKQIISIEFVGPVGEELAQAVATLTL AMVLIYGRSPERMRKRGCSIASLAHVITILGVPSAOIRIDITFVALIISVIGSI DSIVFEDRVENFRKIRLDITIDISLQVTSRTITTSVTLIVVMALFFFGG88I HNFSLALVIGIGETYSIFVALIAVDVGRREHMLPPVQXKIDELP"  
complement(2988.4838)  
gene /gene="HI0240"  
complement(2988.4838)  
CDS /gene="HI0240"  
/note="similar to SP:P19673 GB:S68715 GB:X56175 PID:581230 PID:545175 percent identity: 59.74; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="protein-export membrane protein (secf)"  
/protein\_id="AAC21908.1"  
/db\_xref="GI:1573205"  
/translation="MLNRPYLMKMLAMVIFVAGILYSLPNYGEPAVOISGTRGOE ANTSVGGVODVLTNNLPKTSIVLENGSILAEPTNDDQLADKLAELKNNYTTA LNLAPATFAMLSMGKADPKWGLDRLGCVRFIMVDNNATILVROEQLDLSGELRK EKIQYIAKTEHFGTILITLANSQRAKARIIRQLHPTIDIPDADSINIGLSTA

gene /gene="HI0241"  
complement(4908.5201)  
CDS /gene="HI0241"  
/note="similar to PID:599589 percent identity: 62.37; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAC21909.1"  
/db\_xref="GI:1573206"  
/translation="MEQSPMSTLFPIVIFGLIFPMYTPQAKNKEHKMLSELAK GTEVLTAGVIGIKITKTEGSDSIVIALNDTTEITINNYIVSVLPKSLKSL"  
complement(5309.5530)  
gene /gene="HI0242"  
complement(5309.5530)  
CDS /gene="HI0242"  
/note="similar to GB:L77117 SP:Q58397 PID:1591652 percent identity: 35.71; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAC21910.1"  
/db\_xref="GI:1573207"  
/translation="MKYQNLNLTALRCPIPLSARKKALKNLDKNDMLNLNLSAVEN BSTRPENSVALVEGYVASKRPYIVILKK"  
complement(5527.6045)  
gene /gene="HI0243"  
complement(5527.6045)  
CDS /gene="HI0243"  
/note="hypothetical protein; identified by Genemark; putative"  
/codon\_start=1  
/transl\_table=11  
/product="H. influenzae predicted coding region HI0243"  
/protein\_id="AAC21913.1"  
/db\_xref="GI:1573208"  
/translation="MQILPQQPATTNNEPIEMLYACHSVKRFQRLSTLPYLEKHG YTOAVLNDVEQILSTFNRAAPLHDDDEIDFPQVAPQVOTSIDLEKQHEYLHE MNVALSVQLEELISEORODIDKHLIERFIQYDRHIALEPLFEMGRECTSADILTEM GKMSARQVKE"  
complement(6110.7258)  
gene /gene="HI0244"  
complement(6110.7258)  
CDS /gene="HI0244"  
/note="similar to GB:M63939 SP:P19675 GB:S68715 GB:X56175 PID:147966 percent identity: 81.60; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="trna-guanine transglycosylase (tgc)"  
/protein\_id="AAC21911.1"  
/db\_xref="GI:1573209"  
/translation="MKYELDKTSGNARGRGLVFERPGTFVSVPFAMFVGTGYVKG MPEBRARAGAILLGNTPHMLRPGQVEMRGHGDHDMOHRPILTDSGGQVSL GKURKITEGVAFQNPINGERIFLSPKSMETQYDGSIVWIFDECTYYPATFDYAK KSMEMSLRMAKSRDRFDELGNKNALFGIIOGGVEELKVSLEGLVNIIGFOYVAGG LAVGEKEDMHRILLEYICPOIPADKRPYLMGKEDLVEGVARGIDMDVCWPTPNA RNHGLFVTGIVKIRNAKYRDDTSPLDPECDQCYTCNKYKATLYHIDKCEGELLGALN TTHNLRYVORLMAEIRQATEDDRFDFFVFEFARMGKVPVPQLADKS"  
complement(7602.8693)  
gene /gene="HI0245"  
complement(7602.8693)  
Query Match 5.4%; Score 108.2; DB 1; Length 13818;  
Best Local Similarity 51.4%; Pred. No. 8.7e-14;

Matches 276; Conservative 0; Mismatches 258; Indels 3; Gaps 1;

QY 173 GAAGTGACAACTCGCTCAGATTAAATCTCGAGCGATGTCTTTTGTCTCCGAGGTT 232  
|||  
Db 3590 GAAGCACATACTCTTCTACCTTATTGAATCTGTGCTAATATGACCAATTCAAAT 3531  
|||  
QY 233 CTCAGTGAAGAGCATCTCTTCTGATCTTCTGGAAACAAATGATACACAGGCAATTATC 292  
|||  
Db 3530 GTTGAAGAACCCAAATGGCCCAATCATAGGTGCGCAAAACGATAGACAGGCAATTAAT 3471  
|||  
QY 293 TCAGCATGCTGTGGCTGGCAATGCTTATCTTTGATGAGCGTATATATAGATTTGGA 352  
|||  
Db 3470 GCGAGTCTTTGGGATAGTTGCTGTATTGCCCTTATGTTTGTACTACAAATGTTT 3411  
|||  
QY 353 GCGTGTATCGC---TTGCGAGCTGTTCTTCTGAATCTTTTGTCTATCTGGCAGCTCTA 409  
|||  
Db 3410 GGTGTGATTGCAAGTTTGGCACTTGTATTATATATCGTATTACTTGTGGGATTAATGTCT 3351  
|||  
QY 410 CAGTATTGGAATGCGCACTCACCTTGTGAGCATCTCGTGGGATGTTCTTGTCTGGG 469  
|||  
Db 3350 ATTTTACCGCGCGACACTTTCAATGCGCGGTATGCGGGTATCGTTTAACTTAGGT 3291  
|||  
QY 470 ATGCGCGTAGATCAAAATGTTCTTGTATTTCGAAAGAAATCCGAGAGGAATTTTATTGTCT 529  
|||  
Db 3290 ATGTCAGTAGATCGAATGATATTGTTTGAAGTATTAAAGAGAAATTCGTAATGTT 3231  
|||  
QY 530 CAAAGTCTTAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTAT 589  
|||  
Db 3230 CGTTCAATTCAGCAAGCCATTAATCAAGGTTTATAACGCGCATTTACTTCTATTTTGTAT 3171  
|||  
QY 590 TCTAACTTGACTACAGTATTGGCTCAGCACTTCTTCTTCTCCTAGATACAGGCGCTATT 649  
|||  
Db 3170 GCAAACCTTAAACCAATCTTAAACCGCAATTAATCTATACGCGGTAGGAACAGGCCCAAT 3111  
|||  
QY 650 AAAGGTTTGGCTTTGACATGATTTAGGAATTTCTCTCAATGTTTACGGCTCTT 706  
|||  
Db 3110 CAAGTTTGGCATACGCTTTCACCTTGGTGTGGGATTTCTATGTTTACGCGGATT 3054  
|||

RESULT 29  
AP002996/c 349619 bp DNA linear BCT 15-MAY-2001  
LOCUS Mesorhizobium loti DNA, complete genome, section 3/21.  
DEFINITION AP002996 BA00012  
ACCESSION AP002996.2 GI:14021753  
VERSION  
KEYWORDS  
SOURCE Mesorhizobium loti (strain:MAFF303099) DNA.  
ORGANISM Mesorhizobium loti  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
Phyllobacteriaceae; Mesorhizobium.  
REFERENCE 1 (sites)  
AUTHORS Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,  
Watanabe,A., Igesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,  
Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,  
Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,  
Takeuchi,C., Yamada,M. and Tabata,S.  
Complete genome structure of the nitrogen-fixing symbiotic  
bacterium Mesorhizobium loti  
DNA Res. 7 (6), 331-338 (2000)  
JOURNAL 21082930  
MEDLINE 2 (bases 1 to 349619)  
REFERENCE Kaneko,T.  
AUTHORS Direct Submission  
TITLE Submitted (05-DBC-2000) Takakazu Kaneko, Kazusa DNA Research  
JOURNAL Institute, The First Laboratory for Plant Gene Research, Yana  
1532-3, Kisarazu, Chiba 292-0812, Japan  
(E-mail:kaneko@kazusa.or.jp,  
URL:http://www.kazusa.or.jp/rhizobase/,  
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)  
COMMENT On May 11, 2001 this sequence version replaced gi:11994964.  
FEATURES Location/Qualifiers  
source 1..349619  
/organism="Mesorhizobium loti"

/strain="MAFF303099"  
/db\_xref="taxon:381"  
complement (174..977)  
/gene="ml10869"  
complement (174..977)  
/gene="ml10869"  
/codon\_start=1  
/transl\_table=11  
/product="ion transporter"  
/protein\_id="BAB48365.1"  
/db\_xref="GI:14021754"  
/transl\_table=11  
/transl\_table=11  
/product="DNA polymerase III alpha subunit"  
/protein\_id="BAB48366.1"  
/db\_xref="GI:14021755"  
/transl\_table=11  
/transl\_table=11  
/product="MADRLPRDPLKREAAIAARPEVPARPVHLRVHSAYSLEGA  
LQKLTIVGHAVKDEAPAIATDTNNLFGALFAQAQKVGDIQPIICQIALAFSGNS  
DGRORRQGPMPVLLAATAGYSLVRLVSRVLETPGGEAHLTTMELEGKAD  
GICLSGSRGPIGNALKEDRLAEVRLTLTKLFGDRLYLVDVAGVDRVIEHOT  
IDLAYHELPLVATNEAFSSRDDVEADALJAIAGSVVAVDTRRLSPDNELRQA  
ENQLPADLPEAIDNTVEIALRCSYVYKTRSPILPRPTGGDASDKAAEAAELAK  
QAREGLDARLALHGPPTGYTQYRERLEFELGIIEMKPPGYFLIVADIPWKAQG  
IIPVGRGSGAGSLVAYSTTITDIDPLRFLPERFNPDRVSMPPDFIDFCODRRE  
VIRYQCKYQRQVQIITFTGLQARAVLRVGLQMPYQVDKLSKMPQNPANPV  
KLADAIANEPRAEBAEKPIVOTLLDMAQLKGLYRHAETHAAGIVIGDRPLSELVP  
MYEDPSDMVQFNKMYVEQAGLVKDFELKTLVLETAVKLIERRGIDIDLATIP  
LDDPEYAMLSRGEVGVQVESAGMRKALIGMRPCIEDIIALVALYRRGPNIFPT  
YNARKGEEMASIHPIKDHLVKETQGVIVYQVMOIAQELSGYSLGEADLLRRANG  
KKIARMDKORERFVSGAVRGPQADFIIDLLAKFADYFNKSHAAAYAVVSYG  
AVLKAHYPVEFLAASMTLDWNTDKLDFQDALRLGIEVLAPSVMTSPFPEVGENR  
IYVSMALGVGDAAVEHIVAVRGEKPKNLADFCERVDPKIVGKVFESLIWAGALD  
CGHDAQWAGVERNMGLASLAQANAVSQADIFGASLGAQSOALNLPATDMLAAD  
LRHREQVGVFYLSAHPLDEYKALQMRVQNWAEFSAVVRGASAGRLAGTQVTSKQE  
RTRTKNGKNGVAFSDTSGYEAFLFEALAQYRDLLRAGRSVITVAAEDRPEGVNL  
RINSVQSLBEASRIQALRIFVRNEGPAIIQSQLTORGESQSVSIIVKEEAQGEVE  
IALQGGYRVSPQIASAMRAVPGVVEVELV"  
complement (4811..6823)  
/gene="ml10871"  
complement (4811..6823)  
/gene="ml10871"  
/codon\_start=1  
/transl\_table=11  
/product="propionyl-CoA carboxylase alpha chain precursor"  
/protein\_id="BAB48367.1"  
/db\_xref="GI:14021756"  
/transl\_table=11  
/transl\_table=11  
/product="MFTKILIANRGEIACRVIKTARKMGITATVAVSDADRDRAVHVM  
ADEAVHIGSPAAQSYLVEPKIAACKETGAQVHPGYFLSRAAFCEALEKEGIVF  
IGPKPAIKAMGDKIESKFAANAAYTVPGLGVENADHASKIAGEICVPMWIKAS  
ACGGKGMRIAMNQSEVRGDFPARSEAKSSFGDRAVLEKVFVDRPHIEIOLADAH  
GNALYGERCSIQRRNQAEBAPSPFLDKTRKMGOSVALAKAVDQISAGTVEF  
IVDKQNFYFLENNTRQVHEPVELTVDLVEQMIKVAAGKLSLKQSDIKLNGWA  
VESRLYAEDPYNFLPSIGLTRYRPPPEGQGVVIRNDTGVTEGSEISMTFYDPLMA  
KLCTWAPTRIEADIMSEVLDSPVDGIEHNIIFLAALMQHPRWRGAIATGTIAEY  
PDGFAPVPSREKAVLASIVTAVELLRRDRDLRGLRLPAGHSGVLKRDWVRLGDDY  
LQASILEGMSIPMEIDLSIDGKALTSSDWRPGDLIWRGTGKSVQAQIIPVANG  
LRIAKGMSTATAMLPRIAEERLMPKVPADTSKLLCPCMPGLVVSIAVTEGQEVK  
AGETLAVVSEAMKMNVLRAERDLVSKNAKPGDSLAVDAVIMEFA"  
7023..8150  
/gene="mlr0872"  
7023..8150  
/gene="mlr0872"  
/codon\_start=1  
/transl\_table=11

gene  
CDS  
gene  
CDS  
gene  
CDS  
gene  
CDS



```
/product="glutathione dependent formaldehyde  
dehydrogenase"  
/protein_id="BAB48368.1"  
/db_xref="GI:14021757"  
/translation="MKTRAAVAAVAAAGKPLEIMEVDLDGPREGEVLEIKATGICHTDE  
FTLSCDPEELFPAILGHGAGVAVDGKVTGKGDHVIPLYTECRQPCSLARK  
TNICTAIRATOGGOLMPDSSRSISGDKIENHMGCTGSENFLPEILALAKYNPAP  
PDKICYICGCVTTGICAVINTAKVEQATPVFGIGIGLNTVGLKAGADMTICVD  
LNNDKKAERFGFHTFVNPKEDIDGIVPHLVNMTKRGADQIGADYTPCTCNTKYM  
ROALEASHR3MGKSVVIGVAGAGQEISTRPQVLTGKTWMTAFGGARGKTDVRIVD  
WYMDGKIQIDPMITHLKLEIDINKGFLMHEKGSIRSVVY"  
8113. .8787  
/gene="m1r0874"  
8113. .8787  
/gene="m1r0874"  
/note="hypothetical protein"  
/codon_start=1  
/transl_table=11  
/protein_id="BAB48369.1"  
/db_xref="GI:14021758"  
/translation="MRASRSVASMFTFASHTPTPTDRPATISVONWEKKMAEKLHPRI  
DNGI.PKESASFAGGTLVCACSTRKPVKVKVGOJLAHNNACCTCMKPEGAI.FESVVA  
GTGDDVITVENGDKLVNDASALIQRHACTCGVHMGSPVPRDHA.FKGLTIHFEPRVE  
DGWSPPGFTAFVSSIRESVDPKRMDGIRAOQLRTIGLEPCVLCNPGIMDYIATWTAKK  
SGALPA"  
8818. .9318  
/gene="m1r0875"  
8818. .9318  
/gene="m1r0875"  
/note="hypothetical protein"  
/codon_start=1  
/transl_table=11  
/protein_id="BAB48370.1"  
/db_xref="GI:14021759"  
/translation="MPPAPVRFGEIMPAPTIYVDADAPYKAEKVEKAEKLVVVVYV  
SNGLRSDPMIRNVVSKGADADADIVENKASNDVVTADIPLAIRAVLGAHPTL  
GPTGRPPTPETIGAVAMRDLKQHLRETGSKGNASFAVQDPSRFLGELDIRRAL  
KSATPG"  
9338. .10027  
/gene="m1r0876"  
9338. .10027  
/gene="m1r0876"  
/note="hypothetical protein"  
/codon_start=1  
/transl_table=11  
/protein_id="BAB48371.1"  
/db_xref="GI:14021760"  
/translation="METPAAETPIPRVORHMLFAMSAFCGALAFVALLHAPLAY  
SIGANAPFAVVLIVSOMPKFGKXLSKARADDOVLIPATLVVIVAVVLSFL  
LINOKRGNSEVLEGFALSTPLGFTTHAALAHYAVYMMDDDAVDAREKIKIPUG  
LTFPGDRP.EGMDLVFSAVIGMTAQADTNISITTHRCVVLVHSLISFEFNTIYAA  
AVMLAVSL3BP"  
10125. .11006  
/gene="m1r0877"  
10125. .11006  
/gene="m1r0877"  
/note="hypothetical protein"  
/codon_start=1  
/transl_table=11  
/protein_id="BAB48372.1"  
/db_xref="GI:14021761"  
/translation="MQSVARSQSGATGDADVSSDTLIYQSRRTLTHTHIMASLFF  
MLLSGLQIFNARPOLYIKQSGEGYNNITFIEIGAVNDDAPRGFTETFGHFPDTGVL  
GWSGPPNETNARGPSPWATIPSYVDLGTAVVHFFPFTLITLILVLSVLNQHRL  
RDLAERPDLLKRLPRDVIDHARLKFHHTRENTLOKRAYGVGVLLPLMIITLAMS  
PSMNSVLPFLNEVYGGROTARTIHTFTMLLVGFIFIIHMLLAAGINELRSIITGM  
YRTDPPAHDDKTERSA"  
10975. .11817  
/gene="m1r0879"  
10975. .11817  
/gene="m1r0879"  
/note="hypothetical protein"  
/codon_start=1
```

```
/transl_table=11  
/protein_id="BAB48373.1"  
/db_xref="GI:14021762"  
/translation="MTTRPRGVPMMAKFEISPNFQISRKKLTAAISGASIMLSGC  
DDFSDGLSGLSRFLEMANALTYRAOELLRAOLAPLEFEADIRPORNGVTAP  
DAEDTQKLSANSLRLEMANALTYRAOELLRAOLAPLEFEADIRPORNGVTAP  
WGTPTSLVDADVAPQARYVFPCLDIDSLGDIKYGTIDIDRRHOTLIAY  
GNGKPLPENAPLARYVERQGLKMPRYLKIELVGFANVGGRGVWMDNGDW  
YGGI"  
complement(11866. .12867)  
/gene="m110880"  
complement(11866. .12867)  
/gene="m110880"  
/note="unknown protein"  
/codon_start=1  
/transl_table=11  
/protein_id="BAB48374.1"  
/db_xref="GI:14021763"  
/translation="MGELIMSAPVAVGLTSKNRIAAEDVLMRRDYFADGVTRGEAA  
L.FALDQADKCAEWAPFEVAVTDYIVHGEKPSGVI.SODNADMLVRTISRDGMDSR  
TELELIVHLEKARSSPSKLCYALAEQVHAHYIDKGP.LMLGALVPGI.IAKAEVELL  
RILYAHGSDGNIAITRGEAEVLFRIEBOYAKNDPSMNDL.FVKAIANFVMSAGYE
```

Query Match 5.7%; Score 106.4; DB 1; Length 349619;  
Best Local Similarity 48.2%; Pred. No. 1.5e-13;  
Matches 299; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

QY	150	TCTCAGGGAATTTTACCACCGTGAAAGTGACAAACTCGCTCAGATTTAAATCTGGAG	209
Db	202322	TCTCAGGCAATTTTACCACCGCCAGAGCGCCAAATGATCTGGCCGCTTTGGCAGCCGCTG	202263
QY	210	CGATGCTTTTGTTCGCGAGGCTCTCAGTGAAGAGCGATCTCTTGATCTTGGGAAA	269
Db	202262	CGTGCCGCGAGCGCTGACCGGTATGAGAGACGACGTGGTCCGGCTTTGGCCAGG	202203
QY	270	AACAATGTACACAAGCATATCTCAGACATGCTGTGGCTTGGCAATCTTATTTGTTGA	329
Db	202202	ATTGATTCATGCGCGGAAGAGTGGCGGACATCATCGGTTGATTCCTGTTGGCTTCA	202143
QY	330	TGAGCGTATATTAATGATTTGAGCGGCTCATCGCTTGGGAGCTGTTCTTGAATCTTT	389
Db	202142	TGTTTCGCGCTACCGGCTTCTCGGTTCTCGCCCAACATCGGCTGGCGGTCACGTGG	202083
QY	390	TGCTTATCTGGGACGCTTACAGATTTTGGATGGCGACACCTGTCTGAGACTGCTG	449
Db	202082	CGATGATCGTCGCGCTGTTTCGCTGCTGGGCGAGCGCTGACCTTGGCGATTTGCGG	202023
QY	450	GGATTTGTTCTGTATGGGAGTGGCGTGAATGCAAAATGTTCTGTATTCGAAAGATCC	509
Db	202022	GTAATCGTGAACCATGCGGATGGCGGTGATTCGAAAGTGTGATCTAGAAAGCATCC	201963
QY	510	GAGAGGAATTTTATGTCTCAAAATGCTTAAAAAATCTGAGAAAAGATATACCAAG	569
Db	201962	GTCGAAGAACACGCGGCTGCGCTCGGTGATCCAGGCGATCGACCGGCTTCTCGAAGG	201903
QY	570	CTTTTGGAGCATTTTGTATCTTAACCTGACTACAGATTTGGGCTGACACTTCTTTCT	629
Db	201902	CGCTGGCAGCATGCTGATTCGAAAGTTCAGTCTGATGCGACACCGTGTGCTGTTCT	201843
QY	630	TCCATGATACGAGGCGCTATTAAAGGTTTTCCTTGAACATGATTTAGAAATTTCTTCT	689
Db	201842	ATCTCGGACACCGGCGCGGTGAAGGCTTTCACATACCTACGCGATGGATCTCTGACA	201783
QY	690	CAATGTTTACGCGCTCTTTTCATGATCAATATTTTCTTCATGCTGTGATGAATTAAGACC	749
Db	201782	CGGTTTCACCGCTTTCACCTTACCGGCTGCTGTGATGATGCTGCTGCGCGCCG	201723
QY	750	AACATACACAGTTGATATG	769
Db	201722	GTCGAGAGGAATTCGCGAAG	201703

RESULT 30



AX413017  
LOCUS AX413017 349980 bp DNA linear PAT 14-JUN-2002  
DEFINITION Sequence 8 from Patent WO0228891.  
ACCESSION AX413017  
VERSION AX413017.1 GI:21445475  
KEYWORDS  
SOURCE  
ORGANISM Listeria innocua.  
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
REFERENCE  
AUTHORS Glaser, P. and Kunst, F.  
TITLE Listeria innocua, genome and applications  
JOURNAL Patent: WO 0228891-A 8 11-APR-2002;  
Pasteur Institut (FR)  
FEATURES  
source  
1. .349980  
/organism="Listeria innocua"  
/db\_xref="taxon:1642"  
/note="Original seq 8 (495269) splitted in 2-new seq 8: 1  
to 349980-seq 4026: old seq 8 from 300000 to 495269"  
BASE COUNT 117290 a 59265 c 69878 g 103545 t 2 others  
ORIGIN  
Query Match 5.6%; Score 103.6; DB 6; Length 349980;  
Best Local Similarity 48.8%; Pred. No. 6.5e-13;  
Matches 280; Conservative 0; Mismatches 294; Indels 0; Gaps 0;  
Qy 150 TCTCAGGAAATTTACCCACCGTGAAGTGAGCAAACTCGCCCTCAGATTAAATCTGGAG 209  
Db 190895 TTTCGGGAAGCTTTACACTGAGAGCAAAAGATTAGTCAATGTTAACTCTGGCG 190954  
Qy 210 CGATGCTCTTTGTTCCCGAGGTTCTCAGTGAAGACGATCTCTTGATCTTGGGAAA 269  
Db 190955 CACTTCCTGTAATAAATGACTGAAGTTTACTCTACATCTGTTGGCGCACAGATTGGTCAAG 191014  
Qy 270 AACATGTACAAAGGCAATTAATCTCAGCATGCTGTGGCTTGCAATGCTTATGTTTGA 329  
Db 191015 ATGCATTAACAAGAACTATTTAGTGTGTTATTCGGCGTGTATGCTATCTTTATTTTCA 191074  
Qy 330 TGAGCGGTATATTAGATTGGAGGCGTCATCGCTTCGGGAGCTGTCTTCGTAATCTTT 389  
Db 191075 TGATGCTGTTTACCGTCTACCGGTGTTATCGCATCTATTACATAGTTGCTTACACTT 191134  
Qy 390 TGCTATCTGGCGAGCTCTACAGTATTTGGATGCGCCACTACCTTGTCAGGACTCGTG 449  
Db 191135 APTTAGTTCTATTAATACTAAGCTTACTCAATGCCACACTTACTTTACCAGGATTGCG 191194  
Qy 450 GGATGTTCTTCTGCTATGGGATGCGCGTAGATGCAATGTTCTTGTATTCGAAAGATCC 509  
Db 191195 GCTTGATCTGGGTATAGGTATGGCTGTGACGCAACGTTTATAACATACGAGCGAATAA 191254  
Qy 510 GAGAGGAATTTTATTGTCTCAAGCTTTAAATAATCTGTAGAAAAGGATATACCAAGG 569  
Db 191255 AAGAAGAAATCAAGTTCGGAAGGTCGACAAAGACGACTTTCAGTTCGGTGGAAAAGAG 191314  
Qy 570 CTTTGGAGCCATTTTGTATCTAACTTGAATGAGTATGCGCTCAGCACTCTTTTCT 629  
Db 191315 CATTCGTCGCAATTTAGATGTAATTAACGACACTGATTTGTCGAGTCTTTTCT 191374  
Qy 630 TCCTAGATACAGGCTTATTAAGGGTTGCTTTCACATGATGATTTAGGAATTTCTCTT 689  
Db 191375 ACTTCGGTACAGCTCTATCAAGGGTTTGTACAGTTTCTAATCATTAGTATTCTAGTCA 191434  
Qy 690 CAATGTTTACGGCTCTTTTCATGACTAAATTTT 723  
Db 191435 GCTTCCTAACAGCTGTTGGGTTTCAAGATTCTT 191468  
RESULT 31  
AX417045/c  
LOCUS AX417045 349980 bp DNA linear PAT 14-JUN-2002  
DEFINITION Sequence 4036 from Patent WO0228891.  
ACCESSION AX417045

AX417045.1 GI:21449655  
KEYWORDS Listeria innocua.  
SOURCE Listeria innocua  
ORGANISM Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
REFERENCE  
AUTHORS Glaser, P. and Kunst, F.  
TITLE Listeria innocua, genome and applications  
JOURNAL Patent: WO 0228891-A 4036 11-APR-2002;  
Pasteur Institut (FR)  
FEATURES  
source  
1. .349980  
/organism="Listeria innocua"  
/db\_xref="taxon:1642"  
/note="seq 2058, original length: 3.011.208 replaced  
by-seq 2058: 0.000.001 to 0.349.980-seq 4032: 0.300.001 to  
0.649.980-seq 4033: 0.600.001 to 0.949.980-seq 4034:  
0.900.001 to 1.249.980-seq 4035: 1.200.001 to  
1.549.980-seq 4036: 1.500.001 to 1.849.980-seq 4037:  
1.800.001 to 2.149.980-seq 4038: 2.100.001 to  
2.449.980-seq 4039: 2.400.001 to 2.749.980-seq 4040:  
2.700.001 to 3.049.980-seq 4041: 3.000.001 to 3.011.208"  
BASE COUNT 101566 a 72212 c 58560 g 117642 t  
ORIGIN  
Query Match 5.6%; Score 103.6; DB 6; Length 349980;  
Best Local Similarity 48.8%; Pred. No. 6.5e-13;  
Matches 280; Conservative 0; Mismatches 294; Indels 0; Gaps 0;  
Qy 150 TCTCAGGAAATTTACCCACCGTGAAGTGAGCAAACTCGCCCTCAGATTAAATCTGGAG 209  
Db 68060 TTTCGGGAAGCTTTACAACTGAAGAAGCAAAAGATTAGTCAATGTTAAACTCTGGCG 68001  
Qy 210 CGATGCTCTTTGTTCCCGAGGTTCTCAGTGAAGACGATCTCTTCGATCTTGGGAAA 269  
Db 68000 CACTTCCTGTAATAAATGACTGAAGTTTACTCTACATCTGTTGGCGCACAGTTTGTCAAG 67941  
Qy 270 AACATGTACAAAGGCAATTAATCTCAGCATGCTGTGGCTTGCAATGCTTATGTTTGA 329  
Db 67940 ATGCATTAACAAGAACTATTTAGTGTGTTATTCGGCGTATTGCTATCTTTATTTTCA 67881  
Qy 330 TGAGGATATATTAGATTGGAGGCGTCATCGCTTCGGGAGCTGTCTTCGTAATCTTT 389  
Db 67880 TGAATGCTGTTTACCGTCTACCGGCTGTTATCGCATCTATTACATTTAGTTGCTTACACTT 67821  
Qy 390 TGCTATCTGGCGAGCTCTACAGTATTTGGATGCGCACTCACCTTGTCAGGACTCGCTG 449  
Db 67820 ATTTAGTTCTATTAATTAAGCTTACTCAATGCCACACTTACTTTACCAGGATTGCG 67761  
Qy 450 GGATGTTCTTGTCTATGGGATGCGCGTAGATGCAAAATGTTCTTGTATTCGAAAGAAATCC 509  
Db 67760 GCTTGATCTGGGTATAGGTATGCTGTGACGCAACGTTTATAACATACGAGCGAATAA 67701  
Qy 510 GAGAGGAATTTTATTGTCTCAAGCTTTAAATAATCTGTAGAAAAGGATATACCAAGG 569  
Db 67700 AAGAAGAAATCAAGTTCGGAAGGTCGACAAAGACGACTTTCAGTTCGGTGGAAAAGAG 67641  
Qy 570 CTTTGGAGCCATTTTGTATCTAACTTGAATGAGTATGCGCTCAGCACTCTTTTCT 629  
Db 67640 CATTCGTCGCAATTTAGATGTAATTAACGACACTGATTTGTCGAGTCTTTTCT 67581  
Qy 630 TCCTAGATACAGGCTTATTAAGGGTTGCTTTCACATGATGATTTAGGAATTTCTCTT 689  
Db 67580 ACTTCGGTACAGCTCTATCAAGGGTTTGTACAGTTTCTAATCATTAGTATTCTAGTCA 67521  
Qy 690 CAATGTTTACGGCTCTTTTCATGACTAAATTTT 723  
Db 67520 GCTTCCTAACAGCTGTTGGGTTTCAAGATTCTT 67487  
RESULT 32  
AE006057/c  
LOCUS AE006057 12201 bp DNA linear BCT 08-MAR-2001

[illegible]



gene  
CDS

1897..2598  
/gene="secf"  
1897..2598  
/gene="secf"  
/note="Part of 'sec'-dependent transport complex in prokaryotes; found 5' to the helABOX genes required for cytochrome c biogenesis"  
/codon\_start=1  
/transl\_table=11  
/product="SecF"  
/protein\_id="AB62802.1"  
/db\_xref="GI:2252774"  
/translation="MAFRKLVPKTNPDFRQWATFGAITYMTASVILPIVIGLN FGIIDFGGTTIRTESTTALDVGVRALAELELDVITISEVDFRSDOHVAMITIO MOEDGOGARQOQOQOGLVAKETALVADPALKITISFESVGPVSGELVMTAVSEL AATVIMFYIWRFEWQFALGAVVALVHVLVGLFVQLFEDLTVAALITITGY SINDVVFIDRENLIKYKTMPLRDMNLSTVETLSRTMTGTTLLALVPLWING DVIRGFVPMWVFTGTYSVYVAKNIVLFIIDLNKKEKPSDKFSGAQDAP"

BASE COUNT 496 a 930 c 1015 g 559 t  
ORIGIN

Query Match 5.5%; Score 102.2; DB 1; Length 3000;  
Best Local Similarity 48.4%; Pred. No. 2.2e-12;  
Matches 284; Conservative 0; Mismatches 303; Indels 0; Gaps 0;

QY 150 TCTCAGGAAATTTTACCACCGTGAAGTGACCAACTCGCTCGATTAAATCTGGAG 209  
DB 1262 TCACCGGCGCTTCAGATTCAGAGACCAAGGATCTGGCGCTCTTTTGGCGCGGCG 1321  
QY 210 CGATGCTTTTGTCCGAGAGTTCTCAGTGAAGAGAGATCTTCGATCTTGGGAAA 269  
DB 1322 CCCGCTGCGGGGATGACCTTCCTTGAGAACGACGATGGCGGAGAACCTGGCGCG 1381  
QY 270 AACATGTACAGAGCATTTCTCAGCATGCTGAGCTGGCAATGTTATTTTGA 329  
DB 1382 ACAGCGTCAAGCGGGGATGATGCTGCGATGCGGTTGTCGGGCTGCGCTACA 1441  
QY 330 TGACGATATATATAGATTGGAGCGCTCATGCTTGGGAGCTGTTCTGAATCTTT 389  
DB 1442 TGATGCGCTCCTACGCGCTTTCCGCTTCTTCTGCGGCGCTGTTCAACATTTG 1501  
QY 390 TGCTTATCTGGGCGCTCTACAGATTTGATGCGGCACCTTTCAGAGATCTCGTG 449  
DB 1502 CTTTCATCTTCGCGGTATGCGCGGCGGACGATACCTTCGCGGAGTCGCG 1561  
QY 450 GGATTTGTTCTGATGAGGATGCGCGGTAGATGCAATGTTCTTGAATCGAAGATCC 509  
DB 1562 GGATGCTGCTACCATCGGACCTCGGTGATGCAATGCTGATCTAGAGAGCATGC 1621  
QY 510 GAGAGGAATTTTATGTTCTCAAGTCTTAAAAATCTGTAAGAAAAAGATATACCAAG 569  
DB 1622 GCGAGGAATCGGTCGGGCAATGCGCGCGCGCATGAGCTGGGCTTGAAGAG 1681  
QY 570 CTTTGGAGCATTTTGAATCTTAATCTGACTACAGATTTGCGCTCGACATTTCTTTCT 629  
DB 1682 CGATGCGGCGATATGCAJCGGAACGTGACCTGTTCCCTGCGCGATCTGTTGCG 1741  
QY 630 TCCAGATACAGGCGCTATTAAGGGTTTCTTGAACATGATTTTGAAGATTTTCTTT 689  
DB 1742 TGTGCGGGCGGGCGCGGTGCGCGCTTGTGCGTGAACAGATATGCGCATCGCGCT 1801  
QY 690 CAATGTTTACGGCTCTTTTATGACTAAATTTTCTTCATGCTGTGG 736  
DB 1802 CGATCTTACCGCGATCTG3GTGTCGCGGTGATGATGATGATCTGG 1848

RESULT 34

AX416793/c AX416793 5839 bp DNA linear PAT 14-JUN-2002

LOCUS Sequence 3784 from Patent WO0228891.

DEFINITION AX416793

ACCESSION AX416793.1 GI:21449250

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Pasteur

Institut (FR)

Location/Qualifiers

FEATURES

source

BASE COUNT 1757 a 1256 c 954 g 1872 t

ORIGIN

Query Match 5.5%; Score 102.2; DB 6; Length 5839;  
Best Local Similarity 47.7%; Pred. No. 2.1e-12;  
Matches 299; Conservative 0; Mismatches 328; Indels 0; Gaps 0;

QY 117 TTTTAAACGTCCTCCATTGAAAAATCATGCGATGCTTCAGGAAATTTACCCACCGTGAAG 176  
DB 3730 TTATGTCGCGCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3671  
QY 177 TGAGCAACCTCGCTCGATGATTTAAATCTGAGACGATGCTTGTGTTCCGAGCTTCTCA 236  
DB 3670 CAAAAGTTAGCTGATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3611  
QY 237 GTGAGAGAGATCTCTTCTGATCTTGGAAAAAACAATGATACACAGGATTTACTGAG 296  
DB 3610 ACTGACATCTGTTGGGCTCATTTGGGCAAGATGATGATGATGATGATGATGATGATGAT 3551  
QY 297 CAGCTGTCGCTTGGCAATGCTTATTTGATGATGATGATGATGATGATGATGATGATGAT 356  
DB 3550 GATATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3491  
QY 357 TCAATGCTGCGGAGCTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 416  
DB 3490 TTATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3431  
QY 417 TGATGCGCACCTACCTGTCAGACTCGCTGGATGATGATGATGATGATGATGATGATGAT 476  
DB 3430 TTAACGCAACACTTACTTACAGGATGATGATGATGATGATGATGATGATGATGATGAT 3371  
QY 477 TAGATGCAATGCTTCTGATTTGAAAGAAATCCGAGAGAAATTTATGCTCMAAGTC 536  
DB 3370 TTACGCGGAACGTAAATACCTACAGACGATGATGATGATGATGATGATGATGATGAT 3311  
QY 537 TTAATAAATGATGATAAAGATATACCAAGCTTTGGAGCATTTTGAATTTACTACT 596  
DB 3310 CAAAAGCGCTTTTGAAGTGGCGGTAAAGAGATTTCCGCGATTTTGAAGTGAAT 3251  
QY 597 TGACTACAGATTTGCTCGACACTTCTTTCTTCTGATGATGATGATGATGATGATGATGAT 656  
DB 3250 TGAACACACTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3191  
QY 657 TTGCTTGAACATGATTTTGAAGAAATTTCTTCAATGTTTACGCGCTTTTCAAGACTA 716  
DB 3190 TTGCTGAGGTTTCAATCAATGATTTTACTGACCTTTCAGACGCGCTTGGGAGATCA 3131  
QY 717 AATTTTCTCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 743  
DB 3130 GATTTCTTATGAGCTTATTTAGTAAAA 3104

RESULT 35  
ALS91979/c





```

/Note="CT2261"
163..780
/Genes="rnhB"
/Note="identified by match to PFAM protein family HMM
PF01351"
/codon_start=1
/transl_table=11
/product="ribonuclease HII"
/protein_id="AAW73476.1"
/db_xref="GI:21648308"
/translacion="MLSTDEYPLWESLQVCGIDEAGRGPLAGPVVAAVAIVPRHPR
PTGIFAKLDSDKKLTAEIRDELALAIRESAESWALGVVDAETDIRINILQATLMMNL
AVESLSTGPFVFLVDGRRFRVLPVPYQTVKGSKFVSAASVLAATKTHDELMTTS
AAEYPEYGFVFPYARHVEATARHGRCAIRHQSFLAKLGEK"
800..1189
/Genes="CT2262"
800..1189
/Genes="CT2262"
/Note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAW73477.1"
/db_xref="GI:21648309"
/translacion="MNAPOMLGAGEKIAARHLAAKGYRIVARNYRHRNEIDIAFD
GEALCFEVTKASLGHGKFAESVTRSKQKEIAPAAAGYLASLDDPWITCRFDVIAVL
ALSIDERSIRKYIEHIAKAAFPWVGDKG"
1262..3205
/Genes="gyzB"
/Note="CT2263"
1262..3205
/Genes="gyzB"
/Note="identified by match to PFAM protein family HMM
PF00204"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit B"
/protein_id="AAW73478.1"
/db_xref="GI:21648310"
/translacion="MORTDIQTQNASTEYCATNIQVLDGIEHVRKRPAMVIGDIHSR
GLHLVVEIVDNAIDETLAGVNDYIHVAMPDGSVTVTDHGRGIPVDIHPVKKSALE
LMTVIGAGGKPKDGAVKVSGHLGVGASVUNALSEWCEVYVDGKIPIQTVRRGVP
QGVVEITGTDQRTKVTVPDDEIFKITEFRKDIILDRMRELAFLNSNRKIIVQDRE
GNEIFIEGGLKEFVRFIDNSRLSLKEPIFUSGERDSTMVEIALQNDSYQENIFS
YVNNINHEGHTVHTGFRKATRLNAYAKNDLLKNVKITLGDDEFKGLTAVISVK
VPEPFGQTKTLGNSGTQSVETIVNEQLAFPAESNPGTLLKIIIEKVKSAISREA
ARKAKELTRKSLGSELSGPKLADCSINDPDHCELIVVEGDSAGSAKQGRDRSFOA
ILPLKGLILNVEKARLHMLENEIKTIIILACTSIGEEFSPEKLYRGLIIMTAD
VDGNHITLLTTPFRMRSLIIBAGKVFIAPQPLYLKSGREGQIYAWDEERLSIVEQ
MKKLQKGAANSIQRYKGLGEMNPESQLWSTTMDPAHRSLLQVTVENAMEADQVFSTLM
GDKVEPRDRDFIEKNARYVRLDV"
complement(3194..4510)
/Genes="surA"
/Note="CT2264"
complement(3194..4510)
/Genes="surA"
/Note="identified by match to PFAM protein family HMM
PF00639"
/codon_start=1
/transl_table=11
/product="peptidyl-prolyl cis-trans isomerase SurA"
/protein_id="AAW73479.1"
/db_xref="GI:21648311"
/translacion="MKKLVFLAALMIAMNGFADAASTGLDRIVAIVGNEIILASD
VNEQELMLHQVETKDPOLRKXILLENMINOKIILTAKIDTVKDEKSDVQDAAR
YSSIRAGFPVSAMESRFPNVRNKLKHIEDIRDQOMIEAFRKHFEVTSIDETM
AFYDQKALPEAPETVSVSQIIMPLVSEAAQALDKIKAVQQLEAGGSFATLAR
EYSDPGSGKGBDGLRGELVSPFEAAVSLKPGQISGIVETFRGHIILIDKE
GDIRHTRHIALDRSKTDIPATIALIKSIRKDVLSGKATFAEMAKYSDPASATNG
GLITSGSNPDLEVRPLRKIIDGLSKSGDISOPEKIESDKSAPFYALFNLNSRT
PAHLVTEPHDFAIQELANLHKSQELFNAMIEKLKKEVVVKVMSDI"
complement(4769..7544)
/Genes="CT2265"

```

/note="this region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:AE000513; identified by sequence similarity; putative; protein-export membrane protein secD, putative"

BASE COUNT 1761 a 2213 c 2162 g 1551 t

ORIGIN

Query Match 5.4%; Score 100.6; DB 1; Length 7687;  
Best Local Similarity 51.7%; Pred. No. 4.5e-12;  
Matches 229; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 301 CTGTGGCTTGGCAATGCTATTGTTTTCATCAGCGGTATATATATAGATTGGAGCGGTAT 360  
Db 6186 CTGGGGTTTACGGCGGTGCGTATCTTCATGATGTTTATTATAACAGCGGGTATCGC 6127

QY 361 CGCTTCGGGAGCTCTTCTTCTGAATCTTTTCTGCTTATCTGGGCGCTCTACAGTATTTGGA 420  
Db 6126 CGCCGACCTGGCGCTCGTCTCAACATTATGGTCATCATCGCCGTGCTTCCGGGTTTTCAG 6067

QY 421 TGCGCCACTACCTTGTTCAGACTCGCTGGGATTGTTCTTGTCTATGGGATGGCGGTAGA 480  
Db 6066 CGCCGTACTGACCTTACCCGGTATTGCGGTATCGTGTGACGATGGTATGGCTGTGA 6007

QY 481 TGCAAATGTTCTTCTTATTCGAAAGTCCGAGAGGATTTTATTGTCTCAAAGTCTTAA 540  
Db 6006 CGCCAACTGCTGATTACGAGCGCGTCCGCGAGGAGCTTGCAGCGCAAGAGCCTCG 5947

QY 541 AAAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATTCTAACTTTCAC 600  
Db 5946 CATGGCTATCGACGCGGGCTACACCGTGGTGTTCATGATTCGACTCGACGCTCAC 5887

QY 601 TACAGTATGGCCTCAGACCTTCTTTTCTTCTAGATACAGGGCCTATTAAAGGTTTGC 660  
Db 5886 GACGCTTGGCGCAGCCTTCTGCTCTATATTACGCGGTGGGCGCAATTCAGGATTTCG 5827

QY 661 TTTCACATTGATTTAGCAATTTCTTCAATGTTTACGGCTCTTTTCATGACTAATTT 720  
Db 5826 CGTCACCTCATGATCGGTACTGCGCGAGCTTGTTCACGCGATCGTCTGTCACCGGT 5767

QY 721 TTCTTCATGCTGGGATGAATA 743  
Db 5766 GCTGTTGACCTGCTGCTCGACA 5744

RESULT 37  
AE004160

LOCUS 14261 bp DNA linear BCT 31-JUL-2000  
Vibrio cholerae chromosome I, section 68 of 251 of the complete chromosome.

ACCESSION AE004160 AE003852  
VERSION AE004160.1 GI:9655181  
KEYWORDS  
SOURCE Vibrio cholerae.  
ORGANISM Vibrio cholerae  
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,  
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,  
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,  
Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,  
Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,  
Nierman, W.C. and White, O.  
DNA sequence of both chromosomes of the cholera pathogen Vibrio  
cholerae  
Nature 406 (6795), 477-483 (2000)  
JOURNAL  
MEDLINE 20406833  
PUBMED 10952301  
REFERENCE 2 (bases 1 to 14261)  
AUTHORS Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,  
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,  
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,  
Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,  
Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,



Nierman, W. C., White, O., Salzberg, S. L., Smith, H. O., Colwell, R. R.,  
Mekalanos, J. J., Venter, J. C. and Fraser, C. M.  
Direct Submission  
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712  
JOURNAL  
Medical Center Dr, Rockville, MD 20850, USA

# FEATURES

source

1..14261  
/organism="Vibrio cholerae"  
/strain="V16961"  
/serotype="O1"  
/db\_xref="taxon:666"  
/chromosome="1"  
/note="biotype: El Tor"  
complement(107..616)  
/gene="VC0737"  
complement(107..616)  
/note="similar to GB:LI7309 SP:P39066 PID:348051  
PID:2293317 GB:AL009126; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="acetoin utilization protein Acub, putative"  
/protein\_id="AAF93902.1"  
/db\_xref="GI:9655182"  
/translation="MGKGVDPGSGSLRKKELAMTKVEDMMTRHPHTLRTTLNDA  
KHLMEALIDIRHPITVDANKKLIVSGORDLLAQSLSQASQDSLAPEPTLPEYPM  
TDVTSVAPCGALKEASAIYMQKHIGCLPVAKDVLGITDSQFTTALINLEQES  
EPDELDEC"  
860..1904  
/gene="VC0738"  
860..784  
/gene="VC0738"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF93903.1"  
/db\_xref="GI:9655183"  
/translation="MRTVESGEPDGLQKVTLPEVYHHAFLYLER"  
846..1904  
/gene="VC0739"  
846..1904  
/gene="VC0739"  
/note="similar to GB:M37702 SP:P21516 PID:147485 GB:U00096  
PID:1773089; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="S-adenosylmethionine:tRNA  
ribosyltransferase-isomerase"  
/protein\_id="AAF93904.1"  
/db\_xref="GI:9655184"  
/translation="MOVSDFFHLLPELRLARYPPAERTASRLQLDQNSQGLVDGTEK  
DVLLEP3DLVFNTRVLPARMGRKASGSKGLVERMLDEHTLLAHVRSKSPK  
PGTGLVGENDEPNAWQARHDLPELRFAETVLDLINOIGMPLPYIDRDEEA  
DKERYQTVNCPGAVAPTRGLHFDQDLQAKGVETLAVYTLHAGAGPOGVARE  
NIHDHMAEVEVEPOEVDAITATKARGVAVAGTTSVRSLSAQDALQKTEIK  
PFEGDTEIFIPPGYOYOLVDCLINPHLPSTLLMLVSAPAGYEHTMAAYEHAIVEY  
RFFSYDAMFIRKOMTKA"  
complement(1859..2122)  
/gene="VC0740"  
complement(1859..2122)  
/gene="VC0740"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF93905.1"  
/db\_xref="GI:9655185"  
/translation="MSEKSDHFYPPNRANLRSGQVDSLSIAINKALCGAVAPCEK  
GLHLPALVAVFVPCHLKQPLQVIRIDVALVLCRLINIASP"  
2133..3272  
/gene="VC0741"

CDS

2133..3272  
/gene="VC0741"  
/note="similar to PID:581229 GB:U00096 PID:1773090  
PID:1786607; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="guanine tRNA-ribosyltransferase"  
/protein\_id="AAF93906.1"  
/db\_xref="GI:9655186"  
/translation="MKLKEKLKKGNARRGQLIFERGTVQPAFMPVGYTVKGMT  
PEVEKTAOILLNTTFLMLRPOGVWKKMGDLHDFNMWGPILTDSGQVPSLGD  
IKRTIBESGVFRNPNNGDKIRMDAKSEKIQKDSGLVMTFDECTPYATHDEAKKS  
MEMSLRWAKSRSDHFDKLEPNLNLFGIQQGVYEDLRVSXKGLTETPDGAVGGLA  
VGPEKEDHNRVLEHTCQPLPBDKPRYLMGVGKPEDLVGVGRGIDMPCCVETNARN  
GHLFVYGVVIXIRNAHKTDTPLDPHCDVCYCKNYSXLYHLDRCNCEILGARLNTI  
NHLRYORLMSIRKALIDEDRFDQFVAEFVARRNRPPLQDKA"  
3377..3709  
/gene="VC0742"  
3377..3709  
/gene="VC0742"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF93907.1"  
/db\_xref="GI:9655187"  
/translation="MSLISVAHAGEGAPQGGFEMIIMLAFVIFYPMIYRQAKR  
VKEHNMAAMAKGBEVLTSGLVGRITKIALDNVYITELTNNNEVVIKKDFVAVL  
PKGTLKSL"  
3731..5584  
/gene="VC0743"  
3731..5584  
/gene="VC0743"  
/note="similar to SP:P19673 GB:568715 GB:X56175 PID:581230  
PID:545175; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="protein-export membrane protein SecD"  
/protein\_id="AAF93908.1"  
/db\_xref="GI:9655188"  
/translation="MLNRPYPLMKYMLMTIYVALYALPNYGEEDPAIQTGARGAS  
VMSITDAVTDALNKAQSQKSIALENGSIIVRNDTQVQISARDLISRLGDKIVA  
LNLAPSTPYMLESIGAPMKLGLDLGQVHFLMEVMDAMELVSQDEAPRSDLRD  
EKIRYPAIRPLSDAVEVTLRDAEQVLTLLESKRDMTFTTSBDSGFVLAKEFE  
ARLOEIRNVAECNITILNRNVEIGVAPLVQOGATRIIVELPGVOTARKEITIG  
ATATLEFRVDDKADLAAGAAGAPGSGEIKPDRNGRPVVLKKRVILGSSSIDASS  
ADBYGRPQNISIDSEGNKMGAFSKYKQIKMATYFAVYKSGKSPGKVLITGE  
EVINQATIOSALGRNFRITGIDSPAHNLLALIRAGALIAPISTYBERTIGSMQO  
NIDMGQACIGWGVAVMLFTVLYRKFEMLIANILMANVLIIIVSMIPGATMTLPG  
IAGIVLTGVADVANDVLIIFERIRBELRQGNPQAHQGVANAFSTIADANITTLITA  
ILFAVGTGAIKGFVLTSLIGILTSMTFAIVGRICVNLVGGKIRINKLSI"  
5599..6546  
/gene="VC0744"  
5599..6546  
/gene="VC0744"  
/note="similar to SP:P19674 GB:X56175 PID:581231 GB:U00096  
PID:1773093; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="protein-export membrane protein SecE"  
/protein\_id="AAF93909.1"  
/db\_xref="GI:9655189"  
/translation="MFOILKADKIDFMRSKFAFALSLVMIASIFTLSTKMLNGL  
DPFGGLTIEVGFEPQADRIQISRLSGKDGATVNFSGARVDVNRRLRDMAGGA  
RLAQVLAIAHDVITTLGIFSLQLEVDLITVALTLTVGYSLINTIVFPDIRNF  
RKMRKEPSEAINMSITQTLSTRLTSGTTLFVIALFVGGAMHGFALALLGLITV  
GYSSIVASALAKLIGITREHLPPQVEKEGELDEMP"  
complement(6714..7580)  
/gene="VC0745"  
complement(6714..7580)  
/gene="VC0745"  
/note="similar to GB:M34828 SP:P22783 PID:147899 GB:U00096



Nature 392 (6674), 353-358 (1998)  
98196666  
9537320  
2 (bases 1 to 14567)  
Decker, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L.,  
Graham, D.E., Overbeek, R., Sneed, M.A., Keller, M., Aujay, M.,  
Huber, R., Feldman, R.A., Short, J.M., Olson, G.J. and Swanson, R.V.  
Direct Submission  
Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego,  
CA 92121

Putative indicates no similarity to known proteins  
Hypothetical indicates similarity to a protein of unknown function.

FEATURES  
source  
1. .14567  
/organism="Aquifex aeolicus"  
/strain="VFs"  
/db\_xref="taxon:63363"  
403..912  
/gene="aq\_966"  
403..912  
/gene="aq\_966"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAC07066.1"  
/db\_xref="GI:2983487"  
/transl\_table="MLKDLKQGFVRENHEVDVILGODKEIEFKKIKNATSAVFL  
KGAKILPVNLKIVKDSGLVFGSGSYAMFKHATNVEGLPNLCKPSKEFNLAL  
GLPREKVLISDDIYTDLMGAKELGIGTIFMTTKYKQBELKKNFKPDYVVFHSLKEL  
MREIKKWLVS"  
900..2861  
/gene="osta"  
/note="aq\_967"  
900..2861  
/gene="osta"  
/codon\_start=1  
/transl\_table=11  
/product="organic solvent tolerance protein"  
/protein\_id="AAC07065.1"  
/db\_xref="GI:2983486"  
/transl\_table="WALLILFLLGLSPFSLEKLEIKENELIAGDAEAYKKNY  
IKAKILRLRDKEVYAYEDVYIKNLKTFEIRGSPAYMMKKNGYFLNAKGFKKY  
NFEAKIQEQLDERFKVYATVTTPCLDDKELYLCIFRANITKERALIHNRLKPKFK  
PIFYLPVLIPLGGKRTGLMPTIGSTSYSSYIYQPFWASRDKDMTITADYKDK  
MGLSLERYQADEFESDLITQFSLYKEDKKGDMKMGREYRNRFRITMKYRKKPK  
IGVLSLPDYVEDISFKQEEITKPYTKTYLYEKETKNYLFYQTVRYRDLTEEKN  
AVNLPEASFHUKPYIKGFNFSLDSSFTNFRNYDGSFMRISLEPTVGKIFELFKV  
NYSUTLNNRYPDSEEDKIVSTFRHSPQYFSLSYGGFKFNLEGLYTESPKDY  
EYIYDQDLENNFNQFTLVNLYKNRKLMDYFLNTGYNLIGSYTFPTDGGKIKDK  
KLVTVSVYVVENVRFDQAVDFNLGVFARNVIGTDISWKNFSVGISNSVYKNS  
KLVTTDQIALNLKYSQDFPFCGILSNYDLRLVDTKETYRNVYVGVKGCWSLVLDYKKNY  
YNRREYDQVFLRNLNFKSEIEIPLR"  
complement (2858..3526)  
/gene="cbbE2"  
/note="aq\_968"  
complement (2858..3526)  
/gene="cbbE2"  
/codon\_start=1  
/transl\_table=11  
/product="ribulose-5-phosphate 3-epimerase"  
/protein\_id="AAC07062.1"  
/db\_xref="GI:2983483"  
/transl\_table="MKLLASPLAGDWNCIGEQIQNTIRGGADILHYDVMGDHVFVNI  
TAGYELSHIRKASLPDAHMIENPDKIYIPFVEAGANWISVHIENNYHIRTLOL  
IKELGAKGVVNPQTSLSAIEALYADYILLMSVNPFGSQGKIFERSIERLRIILKE  
NEDRINPCLIBIDGKIKENNAEYVVRAGDVVVVVGSGIFGAKDIEAQTRKLKLSMLS  
AVSV"  
3536..4123  
/gene="cmk"  
/note="aq\_969"  
3536..4123  
/gene="cmk"  
/codon\_start=1



QY	481	TGCAAAATGTTCTTGTATTCGAAGAAATCCGAGAGGAATTTTTATTTGTCTCAAAAGTCTTAA	540
Db	7775	CTCAAAAGTCTCTCATCTTTTGAAGGGTAAAGAGAGAACTGAGACTGGGAAACACCGGTTTC	7834
QY	541	AAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGGATTCTAACTTCAC	600
Db	7835	AAAAGCTATAGAACTTGGGTTTCAGAGAACTTTTAAGTCGGTTTGGGACACACAGGTTAC	7894
QY	601	TACAGTATTGGCCTCAGCACTTCTTTTCTCTCTAGATACAGGGCCATTAAAGGGGTTTGC	660
Db	7895	GCTTCTTGTAGCTCCGTTATACCTTTCAATTTGGGAGCGTCCCGTAAAGGGTTTTC	7954
QY	661	TTTGACATTGATTTTGAAGATTTTCTCT	688
Db	7955	AACGACTCTTGCCTTAGGAACATAGCT	7982
RESULT 39			
AX142455	AX142455	2226 bp	DNA linear PAT 31-MAY-2001
LOCUS	Sequence 1177 from Patent WO0134809.		
DEFINITION	AX142455		
ACCESSION	AX142455		
VERSION	AX142455.1	GI:14282099	
KEYWORDS			
SOURCE	synthetic construct.		
ORGANISM	synthetic construct		
REFERENCE	artificial sequences.		
AUTHORS	1 (bases 1 to 2226)		
TITLE	Kimmerly, W.J.		
JOURNAL	Staphylococcus epidermidis nucleic acids and proteins		
FEATURES	Patent: WO 0134809-A 1177 17-MAY-2001;		
source	GLAXO GROUP LIMITED (GB)		
	Location/Qualifiers		
	1..2226		
	/organism="synthetic construct"		
	/db_xref="taxon:32630"		
	/notes="synthetic nucleic acid sequence"		
BASE COUNT	796 a 320 c 387 g 723 t		
ORIGIN			
Query Match 5.3%; Score 99.6; DB 6; Length 2226;			
Best Local Similarity 48.7%; Pred. No. 8.6e-12;			
Matches 270; Conservative 0; Mismatches 284; Indels 0; Gaps 0;			
QY	173	GAAGTGAGCAAACTCGCCTCAGATTTAAATCTGGAGCGATGTCTTTTGTCCCGAGGTT	232
Db	619	GAGCGAACAATAGCTGAGTTAATAATCGCGCTCATTTACCACTGATTTAAAGAA	678
QY	233	CTCAGTGAAGAGAGATCTCTTTGATCTTGGGAAAAACAATGTACACAGGCATTATC	292
Db	679	ATTTACTCTAACTCTGTTGGTGCAAAATTTGGTCAAGATGCTCTTGATAAGACCATGTTT	738
QY	293	TCAGATGCTGTGGCTTGGCAATGCTTATCTTTTGATGAGCGTATATATAGATTGGA	352
Db	739	GCATCAATTTAGGTATAGCAATTAATTAATTAATTTATGCTTGGTTTCTATCGTTTGCT	798
QY	353	GGCGTCATCGCTTCGGGAGCTGTCTCTGTAATCTTTTGTCTATCTGGGAGCTCTACAG	412
Db	799	GGTTTGTGCAATCATATGCGCTTAACCACTATATTAATTTAACTTAGTCGATTCAT	858
QY	413	TATTTGGANGCGGCACCTCACCTGTGAGGACTCGCTGGGATGTTTCTTGTATGGGGATG	472
Db	859	TTTCATACAGGTGATTAACTCTACCTGGATTGGCGCATTTAGTTTAGGTAGGTATG	918
QY	473	GCCGTAGATGCAATGTCTTGTATTCGAAAGAAATCCGAGAGGAATTTTATTTGTCTCAA	532
Db	919	GCTGTGAGTGCATATATCATTAATGTATGAACGATATAAAGATGAACATAAGAAATTTGGACGC	978
QY	533	AGTCTTAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTTCATTTCT	592
Db	979	ACGCTTAAACAGCGGTATTCAAAGCAATTAAGTTCATTTCTTAACATATTTGATTTCC	1038
QY	593	AACTTGACTACAGTATTTGGCCTTCAGCACTCTTTTCTTCTTAGATACAGGGCCCTATTAAA	652

Db 2625 TTCATATCAGGTATTAATCTACCTGATTTGGCGGATTTAGTTAGTATGATG 2566  
 QY 473 GCCGTAGATGCAATGTTCTTGATTTGCAAAAGATCCAGAGAAATTTTATGTTCA 532  
 Db 2665 GCTGTCCAGTCCAAATATCATATGATATGACGATTAAGATGAACATGAAGATGAGC 2506  
 QY 533 AGCTTTAAAAATCTGTGAAAAAGGATATACCAAGCTTTTGAGGACATTTGATTC 592  
 Db 2605 AGCTTTAAACAGAGTATTAACAAAGCAATAAAGTTATTTACTTATATTTGATTC 2446  
 QY 593 AACTGACATGATTTGGCTCAGACACTTTTCTTCTAGATACAGGCTATTTAA 652  
 Db 2445 AACTTAACAACATGATGCGTGCAGCTGTGCTTTCTTTTGAGAAAGTTCACTCAA 2386  
 QY 653 GGGTTGCTTGAACATTTAGAAATTTTCTTCAATGTTTACGCTTTTCAATG 712  
 Db 2385 GGGTTCGCAACCATGTTACTCTTAGGTATTTAATGATATTTGTAACGCAATTTCTA 2326  
 QY 713 ACTAAATTTTCTT 726  
 Db 2325 TCAAGAGGTTGTT 2312

RESULT 41  
 AX144739/c 3511 bp DNA linear PAT 31-MAY-2001  
 LOCUS AX144739 Sequence 3461 from Patent WO0134809.  
 DEFINITION AX144739  
 ACCESSION AX144739  
 VERSION AX144739.1 GI:14283304  
 KEYWORDS  
 SOURCE synthetic construct.  
 ORGANISM synthetic construct.  
 REFERENCE 1 (bases 1 to 3511)  
 AUTHORS Kimmery, W.J.  
 TITLE Staphylococcus epidermidis nucleic acids and proteins  
 JOURNAL Patent: WO 0134809-A 3461 17-MAY-2001;  
 GLAXO GROUP LIMITED (GB)  
 FEATURES  
 source Location/Qualifiers  
 1.3511  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="synthetic nucleic acid sequence"  
 BASE COUNT 1155 a 610 c 483 g 1263 t  
 ORIGIN

Query Match 5.3%; Score 99.6; DB 6; Length 3511;  
 Best Local Similarity 48.7%; Pred. No. 8.2e-12;  
 Matches 270; Conservative 0; Mismatches 284; Indels 0; Gaps 0;

QY 173 GAAGTGACAACTCGCTAGATTAAATCTGGAGCATGTTCTTTGCCGAGTT 232  
 Db 2865 GAAGGAAACAAATAGCTGAGTTAATATGCGGCTCATTAACGATTGATTTAAAGAA 2806  
 QY 233 CTCAGTGAAGAGCATCTCTTCTGATCTGGGAAAAAACAATGACACAGCATATC 292  
 Db 2805 ATTACTCTAATCTGTTGGTGACAAATTTGGTCAAGTCTCTGATTAAGCAATGTTT 2746  
 QY 293 TCAGATCTGTGGCTTGGCAATGCTTATTTGTTGATGAGGATATTAATAGATTGGA 352  
 Db 2745 GCATCAATTTAGATAGCATTAATTAATTTATGCTTGTGCTTATGCTTGGCT 2686  
 QY 353 GGGCTCATGCTTGGGAGCTGTTCTTGAATCTTTTGTCTATCTGGGACGCTTACAG 412  
 Db 2685 GGTTAGTTGAATCATGCTTAAACCACTTATATTATTTAACTTTAGTGCATTCAT 2626  
 QY 413 TATTGATGGCCCACTACCTTTCAGACATCGCTGGGATTTGTTCTGATGGGAGT 472  
 Db 2625 TTCAATACAGGTATTAATCTTACTGATTTGGGCAATTAAGTTTAAAGTATGAGT 2566  
 QY 473 GCCGTAGATGCAATGTTCTTGATTTGCAAAAGATCCAGAGAAATTTTATGTTCTCA 532  
 Db 2665 GCTGTCCAGTCCAAATATCATATGATATGACGATTAAGATGAACATGAAGATGAGC 2506

QY 533 AGCTTTAAAAATCTGTGAAAAAGGATATACCAAGCTTTTGAGGACATTTTGAATCT 592  
 Db 2605 AGCTTTAAACAGAGTATTAACAAAGCAATAAAGTTATTTACTTATATTTGATTC 2446  
 QY 593 AACTGACATGATTTGGCTCAGACACTTTTCTTCTAGATACAGGCTATTTAA 652  
 Db 2445 AACTTAACAACATGATGCGTGCAGCTGTGCTTTCTTTTGAGAAAGTTCACTCAA 2386  
 QY 653 GGGTTGCTTGAACATTTAGAAATTTTCTTCAATGTTTACGCTTTTCAATG 712  
 Db 2385 GGGTTCGCAACCATGTTACTCTTAGGTATTTAATGATATTTGTAACGCAATTTCTA 2326  
 QY 713 ACTAAATTTTCTT 726  
 Db 2325 TCAAGAGGTTGTT 2312

RESULT 42  
 AF269550/c 3705 bp DNA linear BCT 01-AUG-2000  
 LOCUS AF269550 Staphylococcus epidermidis strain SRI clone step.1010b02 genomic  
 DEFINITION AF269550  
 ACCESSION AF269550  
 VERSION AF269550.1 GI:9623444  
 KEYWORDS  
 SOURCE Staphylococcus epidermidis.  
 ORGANISM Staphylococcus epidermidis.  
 REFERENCE 1 (bases 1 to 3705)  
 AUTHORS Kimmery, W.J., Taylor, J. David., Nelson, A.J., Godlevski, M.M.,  
 Rubino, M.A., Nelson, F.J., Rivers, P.R., Torrella-Miller, I.,  
 Listenebee, S., Ashanti, C., Altschuller, G., Mam, L., Shepherd, N.S.,  
 Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and  
 Furdon, P.J.  
 TITLE Transposon-mediated sequencing of the Staphylococcus epidermidis  
 genome  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3705)  
 AUTHORS Taylor, J. David., Kimmery, W.J., Nelson, A.J., Godlevski, M.M.,  
 Rubino, M.A., Nelson, F.J., Rivers, P.R., Torrella-Miller, I.,  
 Listenebee, S., Ashanti, C., Altschuller, G., Mam, L., Shepherd, N.S.,  
 Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and  
 Furdon, P.J.  
 TITLE Direct Submersion  
 JOURNAL Submitted (22-MAY-2000) Departments of Genomic Sciences and  
 Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore  
 Drive, Research Triangle Park, North Carolina 27709-3398, USA  
 FEATURES  
 source Location/Qualifiers  
 1.3705  
 /organism="Staphylococcus epidermidis"  
 /strain="SRI"  
 /db\_xref="taxon:1282"  
 /clone="step.1010b02"  
 BASE COUNT 1234 a 621 c 509 g 1341 t  
 ORIGIN

Query Match 5.3%; Score 99.6; DB 1; Length 3705;  
 Best Local Similarity 48.7%; Pred. No. 8.2e-12;  
 Matches 270; Conservative 0; Mismatches 284; Indels 0; Gaps 0;

QY 173 GAAGTGACAACTCGCTAGATTAAATCTGGAGCATGTTCTTTGCCGAGTT 232  
 Db 2142 GAAGCGAAACAAATAGCTGATTTAAAGCCGCTCATTAACGATTGATTTAAAGAA 2083  
 QY 233 CTCAGTGAAGAGCATCTCTTCTGATCTTGGAAAAAACAATGATACACAGCATATC 292  
 Db 2082 ATTACTCTAATCTGTGTGATGACATTTGTGCAATGCTCTTGAATAGACATGTT 2023  
 QY 293 TCAGATCTGTGGCTTGGCAATGCTTATTTGTTGATGAGCGATATTAATTAATTTGA 352  
 Db 2022 GCATCAATTTAGGTATAGCATTAATTTATTTATTTATTTGCTGTTCTATCGTTGCT 1963

```

QY 353 GCGTCATCGCTCGGAGCTGTTCTTCTGAATCTTTTGCTTATCTGGCAGCTCTACAG 412
Db 1962 GGTTAGTTGCAATCATGTCCTTAACCACTATATATTTATTTAACTTTAGTCGCATCAAT 1903
QY 413 TATTGGATGCGCCACTCACCTGTCAGGACTCGCTGGGATGTTCTTGCTATGGGATG 472
Db 1902 TTCATATCAGGTGATTAACTCTACCTCGATTGGCGCATTTAGTTAGGTAGGTATG 1843
QY 473 GCGTAGATGCAATGTTCTTGATTCGAAAGATCCGAGAGGAATTTTATATGTCCTCAA 532
Db 1842 GCTGTCGATGCCAATATCATATGATGATGAACGTATTAAAGATGAACCTAAGAAATGGACGC 1783
QY 533 AGTCTTAAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTATCT 592
Db 1782 ACCTTAAACCAAGCGTATTCAAAGCAAAATAAAGTTCAATTTAACTATATTTGATTCC 1723
QY 593 AACTTGACTACAGTATTGGCCTCAGCACTTCTTTCTTCTCCTAGATACAGGCGCTATTAAA 652
Db 1722 ACCTTAAACCAAGCGTATTCAAAGCAAAATAAAGTTCAATTTAACTATATTTGATTCC 1723
QY 593 AACTTGACTACAGTATTGGCCTCAGCACTTCTTTCTTCTCCTAGATACAGGCGCTATTAAA 652
Db 1722 ACCTTAAACCAAGCGTATTCAAAGCAAAATAAAGTTCAATTTAACTATATTTGATTCC 1723
QY 653 GGGTTGCTTTGACATGATTTTAGGAATTTTCTTCTCAATGTTTACGGCTCTTTTCATG 712
Db 1662 GGGTTGCAACCATGTTACTCTTAGGTATTTTAATGATATTTGTACCGCAGTATTCTTA 1603
QY 713 ACTAAATTTTCTT 726
Db 1602 TCAAGAGGGTTGTT 1589

RESULT 43
AX144870/c
LOCUS AX144870 3705 bp DNA linear PAT 31-MAY-2001
DEFINITION Sequence 3592 from Patent WO0134809.
ACCESSION AX144870
VERSION AX144870.1 GI:14283435
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 3705)
AUTHORS Kimmberly, W.J.
TITLE Staphylococcus epidermidis nucleic acids and proteins
JOURNAL Patent: WO 0134809-A 3592 17-MAY-2001;
GLAXO GROUP LIMITED (GB)
FEATURES
    source
        1. .3705
            Location/Qualifiers
                /organism="synthetic construct"
                /db_xref="taxon:32630"
                /note="synthetic nucleic acid sequence"
BASE COUNT 1234 a 621 c 509 g 1341 t
ORIGIN
Query Match 5.3%; Score 99.6; DB 6; Length 3705;
Best Local Similarity 48.7%; Pred. No. 8.2e-12;
Matches 270; Conservative 0; Mismatches 284; Indels 0; Gaps 0;

QY 173 GAAGTGACAACTCGCTCAGATTTAAATCTGGAGCGATGCTTTTGTTCGCGAGTT 232
Db 2142 GAAGCGAAACAATAGCTGAGTTATTAATCGCGCTCATACCACTGATTAAAGAA 2093
QY 233 CTCAGTGAAGAGAGATCTCTTCTGATCTTGGGAAAAACAATGATACAAAGGCAATTATC 292
Db 2082 ATTTACTCTAATCTGTGTGGTGCACAAATTTGGTCAAGATGCTCTTGTATGACCATGTTT 2023
QY 293 TCAGCATGCTGTGCTGGCAATGCTTATCTTTTGTATGACGATATATTAGATTGGA 352
Db 2022 GCATCAATTTAGGTATAGCAATTAATTTATTTATTTATGCTTGTGTTCTATCGTTGCTT 1963
QY 353 GCGTCATCGCTCGGAGCTGTTCTTCTGAATCTTTTGCTTATCTGGCAGCTCTACAG 412
Db 1962 GGTTAGTTGCAATCATGTCCTTAACCACTATATATTTATTTAACTTTAGTCGCATCAAT 1903
QY 413 TATTGGATGCGCCACTCACCTGTCAGGACTCGCTGGGATGTTCTTGCTATGGGATG 472

```

```

Db 1902 TTCAATACAGGTGATTAACTCTACCTGGATTGGCGCATTAGTTTTAGGTGAGGTATG 1843
QY 473 GCGTAGATCAAAATGTTCTTTGATTTCGAAAGAAATCCGAGAGGAATTTTATTGTCCTCAA 532
Db 1842 GCTGTCGATGCCAATATCATATGATGATGAACGTATTAAAGATGAACCTAAGAAATGGACGC 1783
QY 533 AGTCTTAAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTATCT 592
Db 1782 ACCTTAAACCAAGCGTATTCAAAGCAAAATAAAGTTCAATTTAACTATATTTGATTCC 1723
QY 593 AACTTGACTACAGTATTGGCCTCAGCACTTCTTTCTTCTCCTAGATACAGGCGCTATTAAA 652
Db 1722 ACCTTAAACCAAGCGTATTCAAAGCAAAATAAAGTTCAATTTAACTATATTTGATTCC 1723
QY 653 GGGTTGCTTTGACATGATTTTAGGAATTTTCTTCTCAATGTTTACGGCTCTTTTCATG 712
Db 1662 GGGTTGCAACCATGTTACTCTTAGGTATTTTAATGATATTTGTACCGCAGTATTCTTA 1603
QY 713 ACTAAATTTTCTT 726
Db 1602 TCAAGAGGGTTGTT 1589

RESULT 44
AB013081
LOCUS AB013081 4357 bp DNA linear BCT 09-MAY-2002
DEFINITION Thermoanaerobacter tengcongensis strain MB4T, section 108 of 244 of the complete genome.
ACCESSION AB013081
VERSION AB013081.1 GI:20516185
KEYWORDS Thermoanaerobacter tengcongensis.
ORGANISM Thermoanaerobacter tengcongensis
SOURCE Thermoanaerobacter tengcongensis
REFERENCE 1 (bases 1 to 4357)
AUTHORS Bao, Q., Tian, Y., Li, W., Xu, Z., Xuan, Z., Hu, S., Dong, W., Yang, J., Chen, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y., Ling, L., Tan, H., Chen, R., Wang, J., Yu, J. and Yang, H.
TITLE A Complete Sequence of the T. tengcongensis Genome
JOURNAL Genome Res. 12 (5), 689-700 (2002)
MEDLINE 21992816
PUBMED 11997336
REFERENCE 2 (bases 1 to 4357)
AUTHORS Bao, Q., Xu, Z., Hu, S., Dong, W., Chen, Y., Wang, J., Yu, J. and Yang, H.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
REFERENCE 3 (bases 1 to 4357)
AUTHORS Li, W., Xuan, Z., Yang, J., Ling, L. and Chen, R.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
REFERENCE 4 (bases 1 to 4357)
AUTHORS Tian, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y. and Tan, H.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, China
FEATURES
    source
        1. .4357
            Location/Qualifiers
                /organism="Thermoanaerobacter tengcongensis"
                /strain="MB4T"
                /db_xref="taxon:119072"
                /note="putative rho-independent transcription terminator"
terminator 140..159
gene 174..1596
    /gene="AslB3"
    /note="TTE1186"
RBS 174..178
    /gene="AslB3"

```



RESULT 45  
AE003876  
LOCUS  
DEFINITION  
Xylella fastidiosa 9a5c, section 22 of 229 of the complete genome.  
ACCESSION  
AE003876  
VERSION  
AE003876.1  
KEYWORDS  
GI:9105031  
SOURCE  
Xylella fastidiosa 9a5c.  
ORGANISM  
Xylella fastidiosa 9a5c

REFERENCE  
AUTHORS

1 (bases 1 to 11217)  
Simpson, A.J., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alvarado, R., Alves, L.M., Araya, J.E., Baia, G.S., Baptista, C.S., Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R., Bueno, M.R., Camargo, A.A., Camargo, L.E., Carraro, D.M., Carter, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C., Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorri, H., Pacinani, A.P., Ferreira, A.J., Ferreira, V.C., Ferro, J.A., Fraga, J.S., Franca, S.C., Franco, M.C., Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H., Gomes, S.L., Gruber, A., Ho, P.L., Hoheisel, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P., and Marino, C.L.  
The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis  
Nature 406 (6792), 151-157 (2000)  
10910347

TITLE

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

2 (bases 1 to 11217)  
Simpson, A.J.G., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alvarado, R., Alves, L.M.C., Araya, J.E., Baia, G.S., Baptista, C.S., Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R.S., Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M., Carter, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C.R., Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorri, H., Facincani, A.P., Ferreira, A.J.S., Ferreira, V.C., Ferro, J.A., Fraga, J.S., Franca, S.C., Franco, M.C., Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H.S., Gomes, S.L., Gruber, A., Ho, P.L., Hoheisel, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E., Laigret, F., Lambais, M.R., Leite, L.C.C., Lenos, E.G.M., Lenos, M.V.F., Lopes, S.A., Lopes, C.R., Machado, J.A., Machado, W.A., Madeira, A.M.B.N., Madeira, H.M.F., Marino, C.L., Marques, M.V., Martins, E.A.L., Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C., Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, M.A., Nascimento, A.L.T.O., Netto, L.E.S., Nhani Jr., A., Nobrega, F.G., Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C., Palmieri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A.G., Pereira Jr., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodrigues, V., de M. Rosa, A.J., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V., Sawasaki, H.E., da Silva, A.C.R., da Silva, P.R., da Silva, A.M., Silva Jr., W.A., da Silveira, J.F., Silvestri, M.L.Z., Siqueira, W.J., de Souza, A.A., de Souza, A.P., Terenzi, M.F., Truffi, D., Tsai, S.M., Tsuchioka, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S., Vetcore, A.L., Zago, M.A., Zatz, M., Meidanis, J. and Setubal, J.C.  
Direct Submission  
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

TITLE  
JOURNAL

FEATURES  
source

1. .11217  
/organism="Xylella fastidiosa 9a5c"  
/db\_xref="taxon:160492"  
/clone="9a5c"  
229. .1440  
/gene="9a5c"  
229. .1440  
/gene="XF0223"  
229. .1440  
/note="similar to SP|Q54177 (percent identity: 61 %/query alignment coverage: 92.6 %/subject alignment coverage: 99.5 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

gene

CDS

/codon\_start=1  
/transl\_table=11  
/product="queuine tRNA-ribosyltransferase"  
/protein\_id="AAF83036.1"  
/db\_xref="GI:9105032"  
/translation="MGVLVIRDRNGRPLSCACRMSRLOFQLOQATDGAARRGOLSPFCG  
TVPPTLPMVGTGAVKVLPGQLCDLGAIEIILGNTFLHFLRPGLEVIADHGLHSDI-  
RWNPILDTGGGQFVSLAHRKIISQVTFPAATDGAQVFLGPESMKIQKVLHSDI-  
WLFDECTYPATEDVARDMSLRLWAQRSDAHDALDNDALFGLIOGVVPLDRLG  
RSLDGLQALGFDGCGYGGIAGVSESERNVILEYLRPRPADRPRLMGVGRDPLVE  
SVARGVDMFDCVMPITKHARKNQITFGTGVTKIRNACVARDVDIEFGCCGCPACVGGY  
RAYLRHLDRCNEMLASMLGLHNLWYETLMANMRAAITAGTFFAFRRSFYLRGLDL  
PLPEVAGCAG"  
1551. .1913  
/gene="XF0224"  
1551. .1913  
/genes="XF0224"  
/note="similar to SP|P19677 (percent identity: 36 %/query alignment coverage: 90.0 %/subject alignment coverage: 98.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="preprotein translocase YajC subunit"  
/protein\_id="AAF83037.1"  
/db\_xref="GI:9105033"  
/translation="MEITMNLDFRIPVQAQASAPAGAPGGLQMFPLPVLIIAM  
YFSIIRPQMKRQKEYKSMMDHLGCGDEVITAGIAGVSVIGDEFVTEAENVRVY  
QKSAIAKALPKGLSPSAK"  
1980. .3824  
/gene="XF0225"  
1980. .3824  
/genes="XF0225"  
/note="similar to SP|P19673 (percent identity: 44 %/query alignment coverage: 100.0 %/subject alignment coverage: 99.8 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="protein-export membrane protein"  
/protein\_id="AAF83038.1"  
/db\_xref="GI:9105034"  
/translation="MLEIPRMKYVLIFFVFLVLTLYALPNVYOKDP-QITANRGGH  
DDAFQVMTDQAGSIVPKAVKEGSLMVLMSLEMTQTRADYAKVGENYTVL  
NLASTVPLAKLGRPMVLGLDLGVGHAFALQVQKALNCKMFTLARSSENL  
RIGFARVRRPNTIRVTFDDAHLEARMVLFQVGLNYSVDTAEAKRILGATAT  
QIASGVSONLTTLNINELGVAEPIIQOQDDRIIVVELPGVSGDEWVNAVYDE  
LEFRGVDDAYEAVRTGNVPEAKVYIRGTNPVLLNKR-VSGDEWVNAVYDE  
LSPTIAGVFGKNTTGLEAKENSLKLPAGSLAAP-FVEEYVIGSLGAENVE  
RGVAVRSFELFTVFTIYRMFGVITSLLALLFNLLAVMSLFGATMTLPGFAGL  
ALSGLSDVANVLINERIEELRHGMPRSALVAGYDSTILDANLTGLVGVVALY  
AFGTPLKGFALTIIIGIFASMTAITVRSLSALY  
3844. .4809  
/gene="XF0226"  
3844. .4809  
/genes="XF0226"  
/note="similar to GI|1773093 (percent identity: 38 %/query alignment coverage: 87.5 %/subject alignment coverage: 83.1 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"  
/codon\_start=1  
/transl\_table=11  
/product="protein-export membrane protein"  
/protein\_id="AAF83039"  
/db\_xref="GI:9105035"  
/translation="WKLFPADQVQVRLGVGLGNAQVSGVGGNELIRLQVYABH  
NVALEFTGTGLVTRQFAGNPATVQPGFVQVQKLANGVYATFIMLGLFTLY  
NNDVAKNAEAVRKVYDLILTAYMSLREEDLTVLGLLSVMGFAINDIIVF  
IGFRFNSFAVARSILNRSINQTLRTVITAVMPFLSAFALYLVGGRSMEGLAETH  
DRVRENFRSLRVEFLLTIGLVKSVKQDLLPKAKDLESLERRP"  
NAGAVIVVSSV. .5058)  
complement(

gene







GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: December 17, 2002, 11:36:16 ; Search time 46.5 Seconds  
(without alignments)  
10682.974 Million cells updates/sec

Title: US-09-868-987-1  
Perfect score: 3312  
Sequence: 1 atggaactccgcataatgtc.....atgcacaattcagtaataatgc 1864

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlh  
-Q=/cgn2.1/USPTO spool/US09868987/runat 16122002 083005 23518/app\_query.fasta\_1.2055  
-DB=A Geneseq 101002 -OFMT=fastan -SUFFIX=rag -MINMATCH=0 -L=ORCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -LOCALALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=45  
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09868987 @CGN 1.1.45 @runat 16122002 083005 23518 -NCPU=6 -ICPU=3  
-NO XLPXY -NO MWAP -LARGQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGPOP=6 -FCGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A.Geneseq 101002.\*

- 1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2868	86.6	997	20	AAV35184
2	2778	83.9	552	21	AAV95543
3	1593	48.1	1250	20	AAV37780
4	702	21.2	981	20	AAV34476
5	702	21.2	985	20	AAV34352
6	678.5	20.5	737	20	AAV03754
7	647.5	19.6	741	22	AAV82042
8	647.5	19.6	766	23	ABP38844
9	618.5	18.7	754	23	AB448641
10	496	15.0	126	20	AAV37781
11	431	13.0	525	19	AAW98248
12	426	12.9	526	18	AAW55688
13	408.5	12.3	616	22	AA88551
14	408.5	12.3	616	23	AAU91482
15	364	11.0	618	23	AAU91069
16	359.5	10.9	773	22	ABG28914
17	357.5	10.8	713	22	ABG26197
18	351	10.6	323	19	AAW98768
19	349	10.9	140	20	AAV37782
20	342	10.3	325	22	AA88556
21	342	10.3	325	23	AAU91487
22	306.5	9.3	576	22	AAU63631
23	297	9.0	311	23	AAU91065
24	277.5	8.4	614	22	AAU78899
25	277.5	8.4	637	23	AAE15862
26	277.5	8.4	644	22	AAU91576
27	245.5	7.4	279	21	AA805947
28	234	7.1	418	22	AAU63575
29	228	6.9	403	22	AAU91574
30	228	6.9	403	23	AAE15863
31	147.5	4.5	1055	22	AAU36596
32	147	4.4	746	22	AAU96403
33	145	4.4	564	20	AAV35183
34	144	4.3	1034	22	AAU34738
35	140	4.4	178	22	AA882041
36	140	4.2	744	23	AB449954
37	139.5	4.2	507	22	AA896433
38	139	4.2	1098	22	ABG30367
39	136.5	4.1	1007	20	AAV19957
40	136.5	4.1	1032	22	AAU35552
41	136.5	4.1	1036	20	AAV19956
42	136.5	4.1	1051	23	AB40160
43	130.5	3.9	730	22	ABG90442
44	130.5	3.9	730	22	AA878979
45	129	3.9	1040	21	AA815925

ALIGNMENTS

RESULT 1  
AAV35184  
ID AAV35184 standard; Protein; 997 AA.  
XX  
AC AAV35184;  
XX  
DT 13-SEP-1999 (first entry)  
XX  
DE Chlamydia pneumoniae transport polypeptide.  
XX  
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.  
XX  
OS Chlamydia pneumoniae.  
XX  
PN WO9927105-A2.  
XX  
PD 03-JUN-1999.

```

XX PF 20-NOV-1998; 98MO-1B01830.
XX PR 04-NOV-1998; 98US-0107078.
XX PR 21-NOV-1997; 97FR-0014673.
XX PA (GSET ) GENSET.
XX PI Griffiths R;
XX DR WPI; 1999-357842/30.
XX PT Genome sequence of Chlamydia pneumoniae
XX PS Page 1042-1044; Disclosure; 1912pp; English.
CC AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAY91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX SQ Sequence 997 AA;

Alignment Scores:
Pred. No.: 1.3e-315 Length: 997
Score: 2868.00 Matches: 583
Percent Similarity: 99.49% Conservative: 0
Best Local Similarity: 99.49% Mismatches: 3
Query Match: 86.59% Indels: 3
DB: 20 Gaps: 0

US-09-868-987-1 (1-1864) x AAY35184 (1-997)
OY 2 TGGACCTCCGCATATTTGTCAGAGGAGATCAGCGCATGCTGTAATGACAAATATTTCTGCA 61
DB 412 TTTThSerIatATTCyGSGInGluGlyIleSerGlyThrAlaenGlyInTySerAla 431
OY 62 AACGGTATGGCGTATGCGCTGTAGTATGACGCTTATATGTCAGACGCCATTTTAA 121
DB 432 AsnArgIlyTrpArgMetAlaValAlaIleAspGlyTrpMetValSerSerProIleLeu 451
OY 122 AACGTCATTTGAATAATTCAGAGTGTCTCAGGAAATTTACCCACCGGAAGTAGAGC 181
DB 452 AsnAlaProIeuIysAsnH; sAlaSerValSerGlyIysPheTrnHISATGAlValSer 471
OY 182 AAATCCGCTCAGATTTAAATCTGAGCGCATGCTTTTGTCCCGAGGTTCTCAGTGA 241
DB 472 LysIeuAlaSerAspLeuIysSerGlyAlaMetSerPheValProGluValLeuSerGlu 491
OY 242 GAGACGATCTTCTGTATCTTTGGGAAAAACAATGTACACAGGCAATTATCTCAGATGC 301
DB 492 GluThrIleSerSerAspLeuGlyIlyIysGlnCysThrGlnGlyIleIleSerAlaCys 511
OY 302 TGTGGCTTGGCAATGCTTATTTGATGAGCGTATTTTAAATTTGAGGGGCTCATC 361
DB 512 CysGlyIleuAlaMetLeuIleValIleuMetSerValTyTrpAspPheGlyValIle 531
OY 362 GCTTCGGAGAGTGTCTTCTGCAATCTTTGCTATCTGGGAGGCTTACAGATATTTGGAT 421
DB 532 AlaSerGlyAlaValIleuIysAsnIleuIeuIleTrpAlaAlaIeuGlnTyIleuAsp 551
OY 422 GCGGCATCACTTGTAGACACTGCTGGAGATTGTTCTTGTATGGGAGTGGCGGTAGAT 481
DB 552 AlaProIeuThrIeuSerGlyIleuAlaGlyIleValIleuAlaMetGlyMetAlaValAsp 571
OY 482 GCAATGTTCTTGTATTCGAAAGAAATCCGAGAGGAATTTTATGTGCTCAAAGTCTTAAA 541

```

```

DB 572 AlaAsnValIeuValPheGluArgIleArgGluGlnPheIleuIeuSerGlnSerLeuIys 591
OY 542 AAATCTGTAGAAAAAGATATATACCAAGGCTTTTGAGCCATTTTGAATCTTAACTTGACT 601
DB 592 LysSerValGluIysGlyIlyTrnIlysaIlePheGlyAlaIlePheAspSerAsnIleuTrn 611
OY 602 ACAGTATTTGGCCCTCAGACATCTTTCTTCTCTAGATACAGGGCCCTTAAAGGTTTGCT 651
DB 612 ThrValIeuAlaSerAlaIeuIeuPhePheLeuAspThrGlyProIleIysGlyPheAla 631
OY 662 TTGACATTTGATTTTAAAGATTTTCTCTCATGTTTACGCGCTTTTTCATGACTAAATTT 721
DB 632 LeuThrIleuIleuGlyIlePheSerSerMetPheThrAlaIeuPheMetThrIlysphe 651
OY 722 TTCTCATGCTGTGATGATGATAAGACCAACATACACAGTGCATATGATGATAAGTTC 781
DB 652 PhePheMetLeuThrPheIleAsnIlystrGlnHisThrGlnIleuHisMetIleAsnIlysphe 671
OY 782 GTGGGATTAAGCATGATTTTCTTGAAGAGATGCAAAAACTTTGGGCTGTTTCTGGAAGT 841
DB 672 ValGlyIleIysHisAspPheLeuArgGlyCysIlyIysLeuTrpAlaValSerGlySer 691
OY 842 GTTTTCTTTTAAAGTGTGCGTGTCTCGGGTTTGAAGCTGGAATTCGGTTTGGGAATG 901
DB 692 ValPheIleuIeuIlyCysValAlaIleuGlyPheGlyAlaTrpAsnSerValIleuGlyMet 711
OY 902 GATTTTAAAGAGGATATGCTTTTACCTTTAATCAAAAAGAGCATGCAATCAGCATGTT 961
DB 712 AspPheIysGlyIlyTrpAlaPheTrnPheAsnProIyGlnHisGlyIleSerAspVal 731
OY 962 GCTCAATGCTGTGGCAAGTGTGTGCTAAACATACAGAAAGCTGCTTCTTCTTAAGAC 1021
DB 732 AlaGlnMetCArgIlyIysValAlaHisIlyIysLeuGlnAlaIleIysIeuSerSerAsp 751
OY 1022 TTCCGTTATTCAAACATTTGGATCTTCAGAAAAAGATCAAAATCATTTTATGATTAAGCT 1081
DB 752 PheArgIleGlnThrPheGlySerSerGluIysIleIysIleTyPheSerAspIlysaIa 771
OY 1082 TTAAGCTATATCTAAGCA-GATACGAGCCTCTCTCTTAAATTAACGATCATGAGCTGGCG 1140
DB 772 LeuSerTyThrIlysaIleAspThrSerLeuSerProIyIleAsnAspHisGluIeuAla 791
OY 1141 TTAATT-GTGGATTTGTTGTCAGAAACAAGGCTTGATTTCTTACCGAAACTCTAAGCGAA 1199
DB 792 LeuAlaValGlyIleuIeuSerGluTrnGlyLeuAspPheSerThrIuThrIeuAsnGlu 811
OY 1200 ACCCAAAA-TTTTGTCAAAAGGTAAAGCAGCAAACTATCGAAGAAATGCGTTACGGCG 1258
DB 812 ThrGlnAsnPheTrpSerIyIysValSerSerIyIysIeuSerIlyIysMetArgTyrgIlnAla 831
OY 1259 ACCATGGGCTTTTAAAGCTTTGGCAATCATCTTCTATGTAGTTTGGCTTTGAA 1318
DB 832 ThrIleGlyIleuIeuGlyAlaIeuAlaIleIleIleuIeuTyValIleuIeuArgPheGlu 851
OY 1319 TGGCATATATCTTCAAGTCCGATAGCGCTTTAAATTCATGACCTTTTGGTACTCTGCA 1378
DB 852 TrpGlnTyAlaIlePheSerAlaValCysAlaIleuIleHisAspIleuIeuAlaTrnCysAla 871
OY 1379 GTCTTGTTTTATAGACATTTCTTTTGAAGAAATTCAAATAGATTTTGAAGGCATTTGGT 1438
DB 872 ValIeuPheIleAlaHisPhePheIeuIyIysIleGlnIleAspIeuGlnAlaIleGly 891
OY 1439 GCTTTAATGACTGTATTTGGGATATTCATTAACAAATACCTTGTATGATTTTGTATGCTATT 1498
DB 892 AlaIeuMetThrValIleuGlyTySerIleuAsnAsnThrIleuIleIlePheAspArgIle 911
OY 1499 CGTGAAGATGGCAAGCAAGCACTGTTTACCCCTATGATGATGATTTTAAATGATGCGCTT 1558
DB 912 ArgGluAspArgGlnAlaIeuIeuPheTrnProMetHisValIleuValAsnAspAlaIeu 931
OY 1559 CAAAAGCGTTTCAAGCGGACGGTAAATGACAAACAGCTACAGCTTATGATTTGTTAAATG 1618

```

Db 932 GlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMet 951  
Qy 1619 CTTTGTGTTATAGCGCTCTCTGCTTTTAAATTTGCAATTTATGACCATAGGGATT 1678  
Db 952 LeuLeuPheIleGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIle 971  
Qy 1679 CTTCTAGCAACTTATCTCTCTTTATATTGACCACTCTGTTGTTGTTATGTCGCT 1738  
Db 972 LeuLeuGlyThrLeuSerSerLeuTyriLeAlaProProLeuLeuPheMetValArg 991  
Qy 1739 AAAGAAATCGCTCAAAA 1756  
Db 992 LysGluAsnArgSerLys 997

RESULT 2  
AA95543  
ID AA95543 standard; Protein; 552 AA.  
XX AC  
XX AA95543;  
XX 10-OCT-2000 (first entry)  
XX Chlamydia pneumoniae antigen CPN100686 RY 54.  
XX CPN100686 RY 54; antigen; infection; diagnosis; therapy; vaccine;  
XX outer membrane protein.  
XX Chlamydia pneumoniae.  
XX OS  
XX WO200039158-A1.  
XX 06-JUL-2000.  
XX 23-DEC-1999; 99WO-CA01230.  
XX 23-DEC-1998; 98US-0113280.  
XX 23-DEC-1998; 98US-0113281.  
XX 23-DEC-1998; 98US-0113282.  
XX 23-DEC-1998; 98US-0113283.  
XX 23-DEC-1998; 98US-0113284.  
XX 23-DEC-1998; 98US-0113285.  
XX 23-DEC-1998; 98US-0113385.  
XX 28-DEC-1998; 98US-0114056.  
XX 28-DEC-1998; 98US-0114057.  
XX 28-DEC-1998; 98US-0114058.  
XX 28-DEC-1998; 98US-0114059.  
XX 28-DEC-1998; 98US-0114061.  
XX (CONN-) CONNAUGHT LAB LTD.  
XX  
XX Murdin AD, Oomen RP, Wang J;  
XX WPI; 2000-452369/39.  
XX N-PSDB; AAA50030.  
XX  
XX Novel Chlamydia polynucleotides and polypeptides useful for diagnosis,  
XX prevention and treatment of Chlamydia infection in mammals -  
XX Claim 16(a); Page 55-58; 215pp; English.  
XX  
XX The present sequence is that of Chlamydia pneumoniae antigenic  
XX protein CPN100686 RY 54, a putative 93 kDa outer membrane protein.  
XX It is an example of C. pneumoniae antigenic polypeptides of the  
XX invention (see AA95543-55) that are encoded by open reading frames  
XX (see AAA50030-42) identified in the C. pneumoniae genome. The  
XX polypeptides are useful in the diagnosis, treatment and prevention  
XX of Chlamydia infection. They can be prepared by recombinant  
XX methods using transformed unicellular host cells. The  
XX polypeptides, or immunogenic fragments of them, or fusion proteins  
XX that include an additional polypeptide such as a heterologous  
XX signal peptide or a polypeptide having adjuvant activity, are  
XX utilised as vaccines to protect against Chlamydia infection.

CC The polypeptides are also used in diagnostic kits and in methods  
CC of detecting Chlamydia infection.  
XX  
XX SO Sequence 552 AA;  
Alignment Scores:  
Pred. No.: 1.53e-305 Length: 552  
Score: 2778.00 Matches: 552  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 83.88% Indels: 0  
DB: 21 Gaps: 0  
US-09-868-987-1 (1-1864) x AA95543 (1-552)  
Qy 101 ATGGTCAGCAGCCCTATTTTAAACGTCCTCCATTTGAAATCATGCCAGTGTCTCAGGGAAA 160  
Db 1 MetValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
Qy 161 TTTACCCACCGTGAAGTGAGCAAACTCGCCTCAGATTTAAATCTGGAGCGATGCTTTT 220  
Db 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40  
Qy 221 GTTCCCGAGGTTCTCAGTGAAGACGATCTTCTGATCTTGGGAAAAACAATGTACA 280  
Db 41 ValProGluValLeuSerGluThrIleSerSerAspLeuGlyLysGlnCysThr 60  
Qy 281 CAAGGCATTATCTCAGCATCTGTGGCTTGCAATGCTTATTTGATGAGCGTATAT 340  
Db 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTy 80  
Qy 341 TATAGATTGGAGCGTCATCGCTTCGGGAGCTGTTCTTCTGAATCTTTTCTTATCTGG 400  
Db 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrp 100  
Qy 401 GCAGCTCTACAGTATTTGGATGGCCACTACCTTGTCCAGACTCGCTGGGATTGTTCTT 460  
Db 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
Qy 461 GCTATGGGATGGCCGTAGATGCAAAATGTTCTGTTATTTCGAAAGAATCCGAGAGGAATT 520  
Db 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140  
Qy 521 TTATTGTCTCAAGTCTTAAAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCC 580  
Db 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160  
Qy 581 ATTTTGAATCTAACTTGACTACAGTATTTGGCCTCAGACTTCTTTCTTCTCTAGATACA 640  
Db 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180  
Qy 641 GGGCTATTAAAGGGTTTGTCTTGACATTGATTTTAGGAATTTTCTCTCAATGTTTACG 700  
Db 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
Qy 701 GCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAATAAGACCCCAACATACACAG 760  
Db 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisGln 220  
Qy 761 TTGCATATGATGAATAAGTTTCGTGGGATAAAGCATGATTTTGGAGAGGATGCAAAAAA 820  
Db 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240  
Qy 821 CTTTGGGCTGTTCTGGAAGTGTCTTTCTTTAGTTGCGTTCGCTCGGGTTTGGAGCC 880  
Db 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260  
Qy 881 TGGAAATTCGTTTGGGAATGGAATTTTAAAGAGGGGTATGCTTTTACCTTTAAATCCAAA 940  
Db 261 TrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThrPheAsnProLys 280  
Qy 941 GAGCATGGCATCAGCGATGTGCTCAATCGCTGGCAAGTTGTGCATAAATACTACAGAA 1000

Db 281 GluHisGlyIleSerAspValAlaGlnMetArgIlyIysValValHisIlySLeuGlnGlu 300  
 Qy 1001 GCTGGTCTTTCTTTCAGAGATTCCGTAATCAAACTTTGGATCTTCAGAAAGATCAAA 1060  
 Db 301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGlyIlySleIlys 320  
 Qy 1061 ATCTATTTTGTGTAAAGCTTTAAGCTTATCTAAGCAGATAGAGAGCTCTCTCAAAA 1120  
 Db 321 IleTyrPheSerSerPlySAlaLeuSerTyrThrIySglnIleAlaGalaSerLeuLeuIys 340  
 Qy 1121 TTACGATCTAGACCTGAGCGTTATTGTGGATGTGTTCAGAAAACAGGCTAGATTCTC 1180  
 Db 341 LeuThrIleMetSerTrpArgTyrCysGlyIleValValAlaGlnAsnArgProArgPheLeu 360  
 Qy 1181 TACGAAACTCTAAACGAAACGAAATTTTGTCAAGAGTACAGCAAACTATCGAAG 1240  
 Db 361 TyrGlyAsnSerIysAlaGlnAlaIlyPheTrpSerIysValSerSerIysLeuSerIys 380  
 Qy 1241 AAAATGCTTATTCAGCGACCATCGGCTTTTAGAGCTTTGGCAATCATCTGCTCAT 1300  
 Db 381 LysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleIleLeuLeuTyr 400  
 Qy 1301 GTGAGTTGGCGCTTGAATGCGCAATATGCTTTCAGTCCGATGCGCTTAAATCATGAC 1360  
 Db 401 ValSerLeuArgPheGlyTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAsp 420  
 Qy 1361 CTTTGGCTACCTGTGACGCTTGTGTTATAGCACATTTCTTTTGAAGAAATTCAAATA 1420  
 Db 421 LeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuIlySleIleGlnIle 440  
 Qy 1421 GATTGCAAGCGATTTGGCTTTTATAGACTATTTGGGTATTCATTAAACAATCTTG 1480  
 Db 441 AspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnThrIleu 460  
 Qy 1481 ATCATTTTGTATCGATTCCTGAGATGCGCAAGCAACGTTTACCCCTATCATGTT 1540  
 Db 461 IleIlePheAspArgIleAlaGlyIleAspArgGlnAlaAsnLeuPheThrProMetHisVal 480  
 Qy 1541 TTACTTATGATGCGCTTCMAAAGACGTTACGCCGCGATGATGACAAACGATCAACT 1600  
 Db 481 LeuValAsnAspAlaLeuGlnAlaThrPheSerArgTrpValMetThrAlaThrThr 500  
 Qy 1601 CTATCAGTTTGTATATGCTTTTGTATATAGCGCGCTCCCTGCTTAAATTTGCAATT 1660  
 Db 501 LeuSerValLeuLeuMetLeuLeuPheIleGlyIlySerSerValPheAsnPhaIaPhe 520  
 Qy 1661 ATTATGACCATAGGAGATCTCTAGGAACCTTATCGTCTTTTATATGACCACTCTG 1720  
 Db 521 IleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerIlyIleAlaProProLeu 540  
 Qy 1721 TTGTTGTTATGTCCTGTAAGAAATGCTCAAAA 1756  
 Db 541 LeuLeuPheMetValArgIySglnAsnArgSerIys 552  
 RESULT 3  
 AAY37780  
 ID AAY37780 standard; Protein; 1250 AA.  
 XX  
 XX AAY37780;  
 DT 07-OCT-1999 (first entry)  
 XX  
 XX Chlamydia trachomatis outer membrane protein.  
 XX  
 XX Vaccine; eye disease; conjunctivitis; genital disease; perlephacitis;  
 KM paratrachoma; inclusion conjunctivitis; epididymitis; salpingitis;  
 KM nongonococcal urethritis; cervicitis; salpingitis;  
 KM Bartholinitis; pneumonia; venereal lymphogranulomatosis.  
 XX  
 OS Chlamydia trachomatis.  
 XX  
 PN W09928475-A2.  
 XX

PD 10-JUN-1999.  
 XX  
 PF 27-NOV-1998; 98WO-IB01939.  
 XX  
 PR 04-NOV-1998; 98US-0107077.  
 PR 28-NOV-1997; 97ER-0015041.  
 PR 17-DEC-1997; 97ER-0016034.  
 XX  
 PA (GEST) GENSET.  
 XX  
 PI Griffaite R;  
 XX  
 DR MPI; 1999-371125/31.  
 XX  
 PT Genome sequence of Chlamydia trachomatis  
 PS  
 XX Disclosure; Page 1365; 1755pp; English.  
 CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
 CC of Chlamydia trachomatis (see A201425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conjunctivitis, trachoma, nongonococcal urethritis,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC perlephacitis, Bartholinitis; pneumonia; lymphogranulomatosis;  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.  
 XX  
 SQ Sequence 1250 AA;  
 Alignment Scores:  
 Pred. No.: 6.17e-171 Length: 1250  
 Score: 1593.00 Matches: 319  
 Percent Similarity: 75.96% Conservative: 57  
 Best Local Similarity: 64.44% Mismatches: 73  
 Query Match: 48.10% Indels: 46  
 DB: 20 Gaps: 5

US-09-868-987-1 (1-1864) x AAY37780 (1-1250)  
 Qy 2 TGGACCTCCGATATTTGTCAGAGGGATCAGCGGCATGCTATAGCAATATTCGCA 61  
 Db 776 TrpThrSerIysPheCysGlnGlnGlyValSerGlyThrIySasnGlyLeuPheSerGly 795  
 Qy 62 AACCGTGGATGGCGTATGCGTGTAGTATGACGTTATATAGTTCAGACAGCCCTATT 121  
 Db 796 GlyArgGlyTrpArgMetAlaValIleuAspGlyTyrValIleSerAspProValLeu 815  
 Qy 122 AACGTCCCATGMAAATATCGCAGTGTCTCAGGGAATTTACCCACCGTGAAGTACG 181  
 Db 816 AsnValProLeuIysAspHisAlaSerValSerGly\*\*\*PheSerTyrArgGluValHis 835  
 Qy 182 AAACCTGCGCTCAGATTTAAATCTGGAGCGATGCTTTTGTCCCGAGGTTCAGTGA 241  
 Db 836 ArgLeuAlaThrAspLeuIysSerGlyAlaMetSerPheIleProGlnIleLeuSerGlu 855  
 Qy 242 GAGACGATCTCTTCTGATCTTGGAAAAAACAATGTAACACAGCATATCTCAGATGC 301  
 Db 856 GluValValSerProGlnLeuGlySerSerGlnArgValGlnGlyIleLeuSerValVal 875  
 Qy 302 TGTGGCTGGCAATGCTTATATGTTTGTATGAGCGTATATATATATGATTGAGCGCTATC 361  
 Db 876 LeuGlyLeuValValIleuIleValIleuMetSerValTyrTyrArgPheGlyValIle 895  
 Qy 362 GCTTCGGAGACTGCTTCTTGAATCTTTGCTATATCGGAGCTCATCAGATATTTGGAT 421  
 Db 896 AlaSerIleAlaValIleuLeuAsnLeuLeuIleTrpAlaSerMetGlnTyrLeuAsp 915  
 Qy 422 GCGCCACTCACTTGTTCAGACTCGCTGGAGTATTTCTTCTATGCGGAGTGGCGGTAGAT 481  
 Db 916 AlaProLeuThrLeuSerSerGlyLeuAlaGlyIleIleLeuAlaMetGlyMetAlaValAsp 935





XX	WPI; 1999-385613/32.
DR	N-PSDB; AAY91570.
XX	
PT	Antigenic Porphyromonas gingivalis peptides for preventing
PT	gingivitis
XX	
PS	Claim 1; Page 312-314; 588pp; English.
CC	AAY91536 to AAX91801 encode two hundred and sixty six antigenic
CC	Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC	AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC	isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC	activity with a vaccine mechanism of action. The PG polypeptides can be
CC	used as vaccines especially against Porphyromonas gingivalis. Probes can
CC	be used to detect Porphyromonas gingivalis in standard hybridisation
CC	assays. Porphyromonas gingivalis is involved in periodontal disease
CC	especially gingivitis.
XX	
SQ	Sequence 985 AA;
	Alignment Scores:
Pred. No.:	6.32e-70 Length: 985
Score:	702.00 Matches: 193
Percent Similarity:	51.00% Conservative: 112
Best Local Similarity:	32.27% Mismatches: 221
Query Match:	21.20% Indels: 72
DB:	20 Gaps: 17
	US-09-868-987-1 (1-1864) x AAY34352 (1-985)



QY 1703 TATATTGCACACCTCTGTTTGTATTAGTCCGTAAGAAATCGCTCAAAA 1756  
 DB 965 PheValAlaIleThrPro---LeuAlaTyGluIleGlnArgLysLeuAsnLys 981  
 RESULT 6  
 AAY03754  
 ID AAY03754 standard; Protein; 737 AA.  
 XX AAY03754;  
 AC AAY03754;  
 XX  
 DT 10-JUN-1999 (first entry)  
 XX  
 DE B. subtilis secretion factor SecDF.  
 XX Gram-positive microorganism; secretion factor; hormone; growth factor;  
 KW cyclinase; enzyme; protease; cellulase; amylase; carboxylase; lipase;  
 KW reductase; isomerase; epimerase; tautomerase; transferase; kinase;  
 KW phosphatase; SecDF.  
 XX  
 OS Bacillus subtilis.  
 XX  
 PN WO904007-A1.  
 XX  
 PD 28-JAN-1999.  
 XX  
 PF 16-JUL-1998; 98WO-US14786.  
 XX  
 PR 17-JUL-1997; 97EP-0305344.  
 XX  
 PR 16-JUL-1997; 97EP-0305286.  
 XX  
 PA (GENV ) GENENECOR INT BV.  
 PA (GENV ) GENENECOR INT INC.  
 XX  
 PI Quax WJ;  
 XX  
 DR WPI; 1999-132251/11.  
 DR N-PSDB; AAX29477.  
 XX  
 XX Expression vector containing sequence for Gram-positive Sec D, F or  
 PT PF secretion factors - useful for, e.g. increasing secretion of  
 PT heterologous or homologous proteins expressed in host cells  
 XX  
 PS Disclosure; Fig 1A-E; 54pp; English.  
 XX  
 CC The invention relates to an expression vector includes nucleic acid  
 CC encoding a Sec D, F or DF secretion factor from a Gram-positive  
 CC microorganism, under the control of expression signals that cause  
 CC overexpression of the secretion factor in such a microorganism. Cells  
 CC containing the vector are used to secrete secretion factors, heterologous  
 CC or homologous, that is expressed in host cells. Suitable secretion  
 CC factors are hormones, growth factors, cytokines and enzymes (specifically  
 CC a protease, cellulase, amylase, carboxylase, lipase, reductase,  
 CC isomerase, epimerase, tautomerase, transferase, kinase or phosphatase) or  
 CC a ligand, receptor, inhibitor, vaccine or antibody. Expression of these  
 CC secretion factors can overcome the bottle neck created by the limited  
 CC capacity of the secretory system in Gram-positive bacteria, so increases  
 CC production of secreted proteins. The present sequence represents a  
 CC Bacillus subtilis secretion factor SecDF.  
 XX  
 SQ Sequence 737 AA;  
 Alignment Scores:  
 Pred. No.: 2.5e-67 Length: 737  
 Score: 678.50 Matches: 169  
 Percent Similarity: 50.43% Conservative: 125  
 Best Local Similarity: 28.99% Mismatches: 192  
 Query Match: 20.49% Indels: 97  
 DB: Gaps: 15  
 US-09-868-987-1 (1-1864) x AAY03754 (1-737)  
 QY 104 GTTCAGCAGCCTATTATTAAAGTCCCATTTGAAA---AATCATGCCAGTGTCTCAGGGAAA 160

DB 197 ValSerAlaProAsnValSerGlnGlnLeuAsnThrThrAspValLysIleGluGlnHis 216  
 QY 161 TTTTACCACCGGAAAGTAGCAAACTCGCCTCAGATTTTAAATCTGGAGCGATGCTTTT 220  
 DB 217 PheThrAlaGlnGlnAlaLysAspLeuAlaSerIleLeuAsnAlaGlyAlaLeu----- 234  
 QY 221 GTTCCCGAGTTCTCAGTGAAGAG-----ACGATCTCTTGATCTTGGGAAAAA 271  
 DB 235 ---ProValLysLeuThrGlnLysSerThrSerThrSerValGlyAlaGlnPheGlnGln 253  
 QY 272 CAATGTACAAAGGATATATCTACAGATCTGTGGCTTGGCAATGCTTATTTGATG 331  
 DB 254 AlaLeuHisAspThrValPheAlaGlyIleValGlyIleAlaIleIlePheLeuPheMet 273  
 QY 332 AGCGTATATTATAGATTGGAGCGGTCATCGCTTGGAGCGTCTTCTGAACTTTTG 391  
 DB 274 LeuPheTyrrArgLeuProGlyLeuIle-----AlaValIleThrLeuSerVal 290  
 QY 392 CTATCTGGGAGCTCTACAG-----TATTTGATGGCGCACTCACTTGTGAGA 442  
 DB 291 TyrIleTyrrIleThrLeuGlnIlePheAspTrpMetAsnAlaValLeuThrLeuProGly 310  
 QY 443 CTCGCTGGGATTTGTTCTTCTATGGGATGCCGATGATGCAATGCTTGTATTGAA 502  
 DB 311 IleAlaAlaLeuIleLeuGlyValGlyMetAlaValAspAlaAsnIleIleThrTyrrGlu 330  
 QY 503 AGAATCCGAGAGGAATTTTATTTGTCTCAAGCTTTAAAAATCTGTAGAAAAAGATAT 562  
 DB 331 ArgIleLysGlnGlnLeuLysLeuGlyLysSerValArgSerValaLeuArgSerGlyAsn 350  
 QY 563 ACCAAGCTTTTGGAGCCATTTTGTATCTTAACTTGAATGATATGCTTGCAGCACTT 622  
 DB 351 ArgArgSerPheAlaThrIlePheAspAlaAsnIleThrThrIleIleAlaValAl 370  
 QY 623 CTTTCTTCTTACAGAGGCGCTATTTAAAGGCTTGTCTTGCATGATTTTGAATTT 682  
 DB 371 LeuPheIlePheGlyThrSerSerValLysGlyPheAlaIleThrLeuIleLeuSerIle 390  
 QY 683 TTCTCTTCAATGTTTACGCTCTTTTCATGACATAATTTTCTCATGCTG----- 733  
 DB 391 LeuThrSerPheIleThrIleValAlaPheLeuSerArgPheLeuValAlaLeuValGlu 410  
 QY 734 -----TGATGATATAG----- 745  
 DB 411 SerArgTrpLeuAspArgLysGlyTrpPheGlyValAsnLysHisIleMetAsp 430  
 QY 746 -----ACCAACATACACAGTTGCATATGATGAAATTAAGTTCTGGGATTAAG 793  
 DB 431 IleGlnAspThrAspGlnAsnThrGluProHisThrProPheGln-----Lys 446  
 QY 794 CAGATTTTCTTGAGAGATGCCAAAAAATTTGGGCTGTTTGGAGTGTTTTCTTTTA 853  
 DB 447 TrpAspPheThrSerLysArgLysTyrrPhePheIlePheSerSerValThrValAla 466  
 QY 854 GTTGGCTTCTCGCGGTTTGGAGCTTGGATTCGTTTGGGAATGATTTTAAAGA 913  
 DB 467 GlyIleIleIleLeuLeuValPheAlaGlyLeuAsn-----LeuGlyIleAspPheAlaSer 484  
 QY 914 GGG-----TATGCTTATACCTTAAATCCAAAAGCATGCG 949  
 DB 485 GlyAlaArgIleGluValGlnSerAspHisLysLeuThrThrGluGlnValGluLysAsp 504  
 QY 950 ATCAGCGATTTCTCTCAATATGCTGCGCAAGTTGTGCATTAATCAAGAGAGAGTGTCTT 1009  
 DB 505 PheGluSerLeuIleMetAspProAspThrValValLeuSerGlyGlyLysSerAsnIle 524  
 QY 1010 TCT---TCTAGACACTTCGATATCAAACTTGGATCTTCAGAAAAGATCAAAATGTAT 1066  
 DB 525 GlyValAlaArgPheValGlyValProAspLysGluThrIleAlaLysValLysThrTyrr 544  
 QY 1067 TTTAGTATTAAGCTTTTAAAGCTTAACTTAAAGCAGATAGAGAGCTCTCTCTTAAATTAAGC 1126



Db 545 PheLysAspLys----- 548  
 QY 1127 ATCATGAGCTGGCTTATTGTGGGATGTGTCAAAACAGCCCTAGATTCTCTACGGA 1186  
 Db 549 -----TyrGly 550  
 QY 1187 AACTCTAAACGAACGCAAAATTTGGTCAAGGTAAAGCAGCAAACTATCGAAGAAATG 1246  
 Db 551 SerAspProAsnValSerThrValSerProThrValGlyLysGluLeuAlaArgAsnAla 570  
 QY 1247 CQTATATCAGCGCACCCTGGCTTTAGGAGCTTTGGCAATCATCTTGCTCTATGTGAGT 1306  
 Db 571 LeuTyrAlaValAlaIleAlaSerIleGly-----IleIleIleTyrValSer 586  
 QY 1307 TTGCGCTTTGAATGCAATATGCTTTTCAGTGCCTGATGCGCTTTAAATCATGACCTTTTG 1366  
 Db 587 IleArgPheGluTyrLysMetAlaIleAlaIleAlaSerLeuLeuTyrAsp----- 604  
 QY 1367 GCTACCTGTGAGTCTTGTATTATAGCACATTTCTTTTGAAGAAATTCAAATAGATTG 1426  
 Db 605 -----AlaPhePheIleValThrPhePheSerIleThrArgLeuGluValAspVal 621  
 QY 1427 CAAGCATTGGTCTTAAATGACTGTATTGGGGTATTCAATAACAATACTTTGATCAT 1486  
 Db 622 ThrPheIleAlaAlaIleLeuThrIleIleGlyTyrSerIleAsnAspThrIleValThr 641  
 QY 1487 TTTGATCGTATTCTGTAA-----GATCCCAAGCGAACCTGTGTTACCCCTATGCAT 1537  
 Db 642 PheAspArgValArgGluHisMetLysLysArgLysProLysThrPheAlaAspLeuAsn 661  
 QY 1538 GTTTTAGTTAATGATGCCCTTCAAAAGACGTTTCAGCCGACCGTAAATGACACAGCTACA 1597  
 Db 662 HisIleValAsnLeuSerLeuGlnGlnThrPheThrArgSerIleAsnThrValLeuThr 681  
 QY 1598 ACTCTATCAGTTTGTATGATGCTTTTGTATAGCGGCTCCTCTGCTTTTAATTTTGA 1657  
 Db 682 ValValIleValValThrLeuLeuIlePheGlyAlaSerSerIleThrAsnTheser 701  
 QY 1658 TTTATTATGACCATAGGATCTTCTAGGAACCTTTATCGTCTCTTTATATGSCACCACT 1717  
 Db 702 IleAlaLeuValGlyLeuLeuThrGlyValTyrSerSerLeuTyrIleAlaAlaGln 721  
 QY 1718 CTGTTGTTG 1726  
 Db 722 IleTrpLeu 724  
 RESULT 7  
 AAG82042  
 ID AAG82042 standard; Protein; 741 AA.  
 AC AAG82042;  
 AC AAG82042;  
 DT 03-SEP-2001 (first entry)  
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:1178.  
 DX Staphylococcus epidermidis SR1 strain; infection; diagnosis;  
 KW vaccination; endocarditis.  
 XX Staphylococcus epidermidis.  
 OS Staphylococcus epidermidis.  
 PN WO200134809-A2.  
 XX 17-MAY-2001.  
 PD 09-NOV-2000; 2000WO-US30782.  
 XX 09-NOV-1999; 99US-0164258.  
 PR (GLAX ) GLAXO GROUP LTD.  
 PA  
 XX  
 PI Kimmerly WJ;  
 XX

DR WPI; 2001-316495/33.  
 XX N-PSDB; AAH52892.  
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 useful for vaccinating against infections, e.g. endocarditis -  
 XX  
 PS Claim 18; Page 338; 2188pp; English.  
 XX  
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX  
 SQ Sequence 741 AA;  
 Alignment Scores: Length: 741  
 Pred. No.: 8.25e-64 Matches: 165  
 Score: 647.50  
 Percent Similarity: 50.95% Conservative: 129  
 Best Local Similarity: 28.60% Mismatches: 204  
 Query Match: 22 Indels: 79  
 DB: Gaps: 12  
 US-09-868-987-1 (1-1864) x AAG82042 (1-741)  
 QY 89 ATTACGGTTATATGGTCAGCAGCGCCTATTTAAACGTCCCATTTGAAATAATCATGCCAGT 148  
 Db 179 IleSerAlaAlaSerValAspGlnProIleAsnSerSerValGluIleSerGlyGly 198  
 QY 149 GTCTCAGGGAATTTACCCACCGTGAAGTGAGCAAACTCGCTCAGATTTTAAATCTCGA 208  
 Db 199 PheAsnGlyLysLysGlyValGluGluAlaLysGlnIleAlaGluLeuAsnAlaGly 218  
 QY 209 CGCATGTCTTTTGTTCGAGGTCTCAGTGAAGAGACGATCTCTCTGTGATCTTGGGAAA 268  
 Db 219 SerLeuProValAspLeuLysGluIleTyrSerAsnSerValGlyAlaGlnPheGlyGln 238  
 QY 269 AAACAATGTACAAAGGCAATTATCTCAGCATGTGTGGCTTGGCAATGCTTATTGTTTGG 328  
 Db 239 AspAlaLeuAspLysThrMetPheAlaSerIleValGlyIleAlaLeuIleTyrLeuPhe 258  
 QY 329 ATGACGTATATTATAGATTGAGGGCTCATCGCTCGGAGCTGTCTCTGATCTT 388  
 Db 259 MetLeuGlyPheTyrArgLeuProGlyLeuValAlaIleIleAlaLeuThrThrTyrIle 278  
 QY 389 TTGCTTATCTGGCAGCTCTACAGTATTGTGATGCGCCACTCACCTTGTCTCAGGACTCGCT 448  
 Db 279 TyrLeuThrLeuValAlaPheAsnPheIleSerGlyValLeuThrLeuProGlyLeuAla 298  
 QY 449 GGGATTGTTCTGTATGCGGATGCGCGTAGATGCAATATGTTCTTGTATTTCGAAGAATC 508  
 Db 299 AlaLeuValLeuGlyValGlyMetAlaValAspAlaAsnIleIleMetTyrGluArgile 318  
 QY 509 CGAGAGGAATTTTATTGCTCTCAAGTCTTAAATACTGTAGAAAAGGATATACCAAG 568  
 Db 319 LysAspGluLeuArgIleGlyArgThrLeuLysGlnAlaTyrSerLysAlaAsnLysSer 338

```

QY 569 GCTTTGAGCCATTTTGAATCTGACTATGAGTATGGCCAGCACTCTTTC 628
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 SerPheLeuThrIlePheAspSerAsnLeuThrValIleAlaIleValPhe 358
QY 629 TTCTGATGATACAGGCGCTTAAAGGTTTCTTGACATTTTGAATTTTCTC 668
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 PhePheGlyGluSerSerValLysGlyPheAlaThrMetLeuLeuGlyIleLeuMet 378
QY 689 TCATGTTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTG----- 733
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 379 IlePheValThrAlaValPheLeuSerArgGlyLeuLeuSerLeuValSerSerAsn 398
QY 734 -----TGATGATATAGACCCACATACAGCTGATGATGATGAT 775
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 399 PhePheIleYsglnIleTyrTrpLeuPheGlyValIleYsglySAspArgHisAspIleAsn 418
QY 776 AAGTCGTGGGATTAAGCATGAT-----TTCTTGAGA 808
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 GluGlyLysAspVal---HisAspLeuLysThrSerTyrGluArgLeuAsnPheValLys 437
QY 809 GGATGCAAAAACCTTGGCGTCTTTCTGGAAGTCTTTTCTTTAGTTCGCTGCTC 868
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 LeuAlaLysProLeuIleSerLeuSerIleLeuIleValIleIleGlyLeuIleIle 457
QY 869 GGCTTGGAGCCCTGGAATTCGTTTGGGAATGATTTTAAAGAGGCTATGCTTACC 928
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 458 SerIlePheLysLeuAsn-----LeuGlyIleAspPheSerSerGlyThrArgAlaAsp 475
QY 929 TTTAATCCAAAAGAGATGCGATCAGCATGCTGCTCAAAATGCGTGGCAAAAGTTGTCAT 988
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 IleGlnSerIysAsn-----AlaIleThrGlnAlaGlnValGluLys 489
QY 989 AAACATCAGAGAGAGTGTCTTTCTTCTAGAGACTCCGCTATCCAAATTTGATGCTCTCA 1048
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 490 ThrValIysSerValGlyLeuGluProAsp-----GlnIleGlnIleAsnGlySerGly 507
QY 1049 GAAAGATCAAAATCTATTAGATGAAGCTTAAAGCTATAGACATGACATGCGAGCC 1108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 508 AsnLys-----AsnAlaThrValGlnPheLysLys 517
QY 1109 TCTCTCTAAATTAACGATCATGAGCTGGCGTATTTGGGATTTGTTCAGAAACAG 1168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 518 AspLeuSerArg----- 521
QY 1169 CCTAGATTCTCTACGGAATCTTAAACGAAACGCAAAATTTGCTCAAG----- 1219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 522 -----GluGluAspAsnLysLeuSerAlaLysValLysSerGluPheGlyAsp 537
QY 1220 -----GTAAGACGAAACTATCGAAGAAATAGCGTTTACAGCGACCC 1261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 AsnProGlnIleAsnThrValSerProLeuIleGlyGlnGluLeuAlaLysAsnAlaVal 557
QY 1262 ATCGGCGCTTTAGAGCTTTGGCAATCATCTTCTCATATGATGATTTGGCTTTGAATGG 1321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 558 ThrAlaLeuIleLeuAlaSerIleGlyIleIleIleIleThrValSerLeuArgPheGluTrp 577
QY 1322 CAATATCTTTCAGTGGCGATGCGCTTAAATCATGACTTTTGGCTACTGTGCACTC 1381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 578 ArgMetGlyLeuSerSerValLeuAlaLeuLeuHisAspValPhe----- 592
QY 1382 TTGTTTATAGACATTTCTTTTGAAGAAATTCAAATATGTTGCAAGCATGCTGCT 1441
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 593 IleIleIleAlaIlePheSerLeuPheArgLeuGluValAspLeuThrPheIleAlaAla 612
QY 1442 TTAATGACTGATTTGGGATTTATTAACAATATGATGATTTTGAATGCTATTCGT 1501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 613 ValIleuThrIleValGlyTyrSerIleAsnAspThrIleValThrPheAspArgValArg 632
QY 1502 GAAGAT---CGCCAGCGAAGCTGTTTACC-----CCTATGATGTTTATGTTAAT 1549
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 633 GluAsnLeuHisLysValLysValIleThrHisThrAspGlnIleAspAspIleValAlaAsn 652
QY 1550 GATGCCCTTCAAAAGACCTTACGCCGACGCTATGACAAAGCTATCAACTCTATCAGTT 1609

```

```

Db 653 ArgSerIleArgGlnThrThrThrArgSerIleAsnThrValLeuThrValValValVal 672
QY 1610 TTGTTATATCTTTTGTATTATAGCGGCTCTGCTGCTTAAATTTTGAATTTATGACC 1669
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 673 ValValAlaIleLeuIleLeuGlyAlaProThrIlePheAsnPheSerLeuAlaLeu 692
QY 1670 ATAGGATCTTCTTCAAGCACTTATGCTCTTATATGACACCACTCTG 1720
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 693 IleGlyLeuLeuSerGlyValPheSerSerIlePheIleAlaValProLeu 709

RESULT 8
ABP38844
ID ABP38844 standard; Protein: 766 AA.
XX
AC ABP38844;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3689.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064364P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI, 2002-381255/41.
DR N-PSDB; ABN91389.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
PS Disclosure; SEQ ID 3689; 267bp; English.
XX
ABN90538 to ABN93174 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 766 AA;
XX

Alignment Scores:
Pred. No.: 8 4e-64 Length: 766
Score: 647.50 Matches: 165
Percent Similarity: 50.95% Conservative: 129
Best Local Similarity: 28.60% Mismatches: 204
Query Match: 19.55% Gaps: 79
DB: 23 Indels: 12

US-09-868-987-1 (1-1864) x ABP38844 (1-766)
QY 89 ATTGACGGTTATATGCTGACAGCCCTATTTTAAACGTCCTCATGAAATCATGCGCAGT 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 204 IleSerAlaAlaSerValAspGlnProIleAsnSerSerValGluIleSerGly 223  
 Qy 149 GTCTCAGGGAATTTACCCACCGTGAAGTAGAGCAAACTCGCTCAGATTTAAATCTGGA 208  
 Db 224 PheAsnGlyLysLysGlyValGluGluAlaLysGlnIleAlaGluLeuLeuAsnAlaGly 243  
 Qy 209 GCGATGTCTTTTGTCCGAGGCTTCTCAGTGAAGAGAGCATCTCTTCATCTGGGAAA 268  
 Db 244 SerLeuProValAspLeuLysGluIleTyrSerAsnSerValGlyAlaGlnPheGlyGln 263  
 Qy 269 AAACAATGTACACAGGATATCTCAGCATGCTGTGCTGGCAATCGTATTCTTTTG 328  
 Db 264 AspAlaLeuAspLysThrMetPheAlaSerIleValGlyIleAlaLeuIleTyrLeuPhe 283  
 Qy 329 ATGACGCTATATTATAGATTGGAGCGCTCATCGCTTCGGAGCTGTCTCTCGAATCTT 388  
 Db 284 MetLeuGlyPheTyrArgLeuProGlyLeuValAlaIleIleAlaLeuThrTyrIle 303  
 Qy 389 TTGCTTATCTGGCAGCTCTACAGATTATGGATGGCCACTCACCTCTCAGGACTCGCT 448  
 Db 304 TyrLeuThrLeuValAlaPheAsnPheIleSerGlyValLeuThrLeuProGlyLeuAla 323  
 Qy 449 GGGATTGTCTTGTATGGGATGGCGCTAGATGCAAAATGTTCTTGATTCGAAGAATC 508  
 Db 324 AlaLeuValLeuGlyValGlyMetAlaValAspAlaAsnIleIleMetTyrGluArgIle 343  
 Qy 509 CGAGAGGAATTTTATGTCTCAAAAGTCTTAAAAATCTGTAGAAAAAGGATATACCAAG 568  
 Db 344 LysAspGluLeuArgIleGlyArgThrLeuLysGlnAlaTyrSerLysAlaAsnLysSer 363  
 Qy 569 GCTTTTGAGCCATTGTGATCTAACTGACTACAGTATTGGCTCAGCCTCTCTTTTC 628  
 Db 364 SerPheLeuThrIlePheAspSerAsnLeuThrValIleAlaAlaValLeuPhe 383  
 Qy 629 TTCTTAGATACAGGCCCTATTAAAGGTTTGTCTTGACATTTAGGAATTTTCTCT 688  
 Db 384 PhePheGlyGluSerSerValLysGlyPheAlaThrMetLeuLeuLeuGlyIleLeuMet 403  
 Qy 689 TCAATGTTTACGCTCTTTTCATGACTAAATTTTCTTCATGCTG----- 733  
 Db 404 IlePheValThrAlaValPheLeuSerArgGlyLeuLeuSerLeuValSerSerAsn 423  
 Qy 734 -----TGGATGAATAGACCCCAACATACACATGTCATATGATGAT 775  
 Db 424 PhePheLysLysGlnTyrTrpLeuPheGlyValLysLysMetAspArgHisAspIleAsn 443  
 Qy 776 AAGTTCGTGGGATAAAGCATGAT-----TTCTTGAGA 808  
 Db 444 GluGlyLysAspVal--HisAspLeuLysThrSerTyrGluArgLeuAsnPheValLys 462  
 Qy 809 GGATGCAAAAACTTTGGCGTGTCTCGAAGTGTCTTTTCTTTTAGGTGGCTGCTCTC 868  
 Db 463 LeuAlaLysProLeuIleSerLeuSerIleLeuIleValIleIleGlyLeuIleIle 482  
 Qy 869 GGGTTTGAGCTGGAATCCGTTTGGGAATGGATTTTAAAGGAGGATGTCCTTTACC 928  
 Db 483 SerIlePheLysLeuAsn-----LeuGlyIleAspPheSerSerGlyThrArgAlaAsp 500  
 Qy 929 TTTAATCCAAAAGAGCATGGCATCAGCATGTTGTCTCAAAATCGTGCAAGTCTGTCAT 988  
 Db 501 IleGlnSerLysAsn-----AlaIleThrGlnAlaGlnValGluLys 514  
 Qy 989 AAATCAGGAAGCTGGTCTTTCTTCTAGACATTCCTGATTTCAACATTTGGATCTTCA 1048  
 Db 515 ThrValLysSerValGlyLeuGluProAsp-----GlnIleGlnIleAsnGlySerGly 532  
 Qy 1049 GNAAGATCAAAATCTATTATTAGTGATAAAGCTTTAAGCTATACTAAGCAGATACAGACC 1108  
 Db 533 AsnLys-----AsnAlaThrValGlnPheLysLys 542  
 Qy 1109 TCTCTCTAAATTAACGATCATGAGTGGCGTATTATGCGGATTTGTTGTCAGAACAGG 1168  
 Db 543 AspLeuSerArg----- 546

Qy 1169 CCTAGATTTCTCTACGAAACTCTAAACGAAACCCAAAAATTTTGGTCAAAG----- 1219  
 Db 547 -----GluGluAspAsnLysLeuSerAlaLysValLysSerGluPheGlyAsp 562  
 Qy 1220 -----GTAAGCAGCAAACTATCGAAGAAATCGGTATTCACGCGACC 1261  
 Db 563 AsnProGlnIleAsnThrValSerProLeuIleGlyGlnGluLeuAlaLysAsnAlaVal 582  
 Qy 1262 ATCGGGCTTTTAGGAGCTTTGGCAATCATCTGCTCTATGTGAGTTTGGCTTTGAATGG 1321  
 Db 583 ThrAlaLeuIleLeuAlaSerIleGlyIleIleIleTyrValSerLeuArgPheGluTrp 602  
 Qy 1322 CAATATGCTTTCATGCGCTATGCGCTTTAATTCATGACCTTTTGGCTACCTGTCGAGTC 1381  
 Db 603 ArgMetGlyLeuSerSerValLeuAlaLeuLeuHisAspValPhe----- 617  
 Qy 1382 TTGTTTATAGCACATTTCTTTTGAAGAAATTCAAATAGATTGCAAGCCATTGGTGT 1441  
 Db 618 IleIleIleAlaIlePheSerLeuPheArgLeuGluValAspLeuThrPheIleAlaAla 637  
 Qy 1442 TTAATGACTGTATTGGGTATTTCATTAACAATACTTTGATCATTTTTGATCGTATTTCGT 1501  
 Db 638 ValLeuThrIleValGlyTyrSerIleAsnAspThrIleValThrPheAspArgValArg 657  
 Qy 1502 GAAGAT---CGCCAAAGCAACCTGTTTACC-----CCTATGCATGTTTATAGTTAAT 1549  
 Db 658 GluAsnLeuHisLysValLysValIleThrHisThrAspGlnIleAspAspIleValAsn 677  
 Qy 1550 GATCCCTTCAAAAGAGCTTCAGCGCAGCGTAAATGACACAGCTACAACTCTATCAGTT 1609  
 Db 678 ArgSerIleArgGlnThrMetThrArgSerIleAsnThrValLeuThrValValVal 697  
 Qy 1610 TTGTTAATGCTTTTGTATTAGCGGCTCTCTCTCTTTAATTTTGCATTTTATATGACC 1669  
 Db 698 ValValAlaIleLeuIleLeuGlyAlaProThrIlePheAsnPheSerLeuAlaLeuLeu 717  
 Qy 1670 ATAGGGATTTCTTAGGAACCTTATTCGTCTCTCTTTATATTGCACCCTCTG 1720  
 Db 718 IleGlyLeuLeuSerGlyValPheSerIlePheIleAlaValProLeu 734  
 RESULT 9  
 ABB48641  
 ID ABB48641 standard; Protein; 754 AA.  
 XX  
 AC ABB48641;  
 XX  
 DT 05-FEB-2002 (first entry)  
 XX  
 DE Listeria monocytogenes protein #1345.  
 XX  
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KM vitamin B12; bacterial infection; disease.  
 XX  
 OS Listeria monocytogenes.  
 XX  
 PN WO200177335-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 11-APR-2001; 2001WO-FR01118.  
 XX  
 PR 11-APR-2000; 2000PR-0004629.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Feihl H, Dehoux P;  
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;



Db 664 AlaValAsnLysAlaLeuArgGlnThrPheThrArgSerIleAsnThrIleLeuThrVal 683  
 Qy 1601 CTACGAGTTTGTAAATGCTTTGTTTATAGCGGCTCTCTGCTTTAAATTTTGCATTT 1660  
 Db 684 IlePheThrValLeuAlaLeuValLeuPheGlySerGluSerIleLeuAsnPheSerIle 703  
 Qy 1661 ATTATGACCATAGGATCTTCTAGGAACCTTTATGCTCTCTTTATATATGACCACTCTG 1720  
 Db 704 AlaLeuLeuValGlyLeuValSerSerValPheSerSerIlePheMetAlaMetGlnLeu 723  
 Qy 1721 TTGTTGTTTATGTCGTAAGAAATCGCTCAAA 1756  
 Db 724 TrpTyrValPheLysAlaArgGlnLeuArgLysLys 735  
 RESULT 10  
 AAY37781  
 ID AAY37781 standard; Protein; 126 AA.  
 XX  
 AC AAY37781;  
 XX  
 DT 07-OCT-1999 (first entry)  
 XX  
 DE Chlamydia trachomatis outer membrane protein.  
 XX  
 KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.  
 XX  
 OS Chlamydia trachomatis.  
 XX  
 FN WO9928475-A2.  
 XX  
 PD 10-JUN-1999.  
 XX  
 PF 27-NOV-1998; 98WO-IB01939.  
 XX  
 PR 04-NOV-1998; 98US-0107077.  
 PR 28-NOV-1997; 97FR-0015041.  
 PR 17-DEC-1997; 97FR-0016034.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Griffais R;  
 XX  
 DR WPI; 1999-371125/31.  
 XX  
 PT Genome sequence of Chlamydia trachomatis  
 XX  
 PS Disclosure; Page 1365-1368; 1755pp; English.  
 XX  
 CC AAY3754-Y37949 are encoded by open reading frames (ORFs) of the genome  
 CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conventional trachoma, nonendemic trachoma,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC perithepatitis, bartholinitis; pneumopathy in breast feeding infants;  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.  
 XX  
 SQ Sequence 126 AA;  
 Alignment Scores:  
 Pred. No.: 4.87e-47 Length: 126  
 Score: 496.00 Matches: 97  
 Percent Similarity: 91.22% Conservative: 13  
 Best Local Similarity: 82.20% Mismatches: 8  
 Query Match: 14.98% Indels: 0  
 DB: 20 Gaps: 0

US-09-868-987-1 (1-1864) x AAY37781 (1-126)  
 Qy 1400 TTTTGAAGAAAAATTCAATAGATTTCAGAGCCATTGGTCTTAATGACTGTATTGGG 1459  
 Db 6 PheCysArgGluLeuGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGly 25  
 Qy 1460 TATTCAATAAACAATACTTTGATCATTTTGTATGCTGATTGCTGAAGATCGCAAGCGAAC 1519  
 Db 26 TyrSerLeuAsnAsnThrLeuIlePheAspArgIleArgGluAspArgGluLys 45  
 Qy 1520 CTGTTTACCCCTATGCATGTTTGTAGTTAATGATGCCCTTCAAAAGACGTTTCAGCGCGACG 1579  
 Db 46 LeuPheThrProMetProIleLeuIleAsnAspAlaLeuGlnLysThrLeuGlyArgThr 65  
 Qy 1580 GTAATGACACAGCTACAACCTCTATCAGTTTGTGTTTGTATGCTTTTATAGGCGGCTCC 1639  
 Db 66 ValMetThrThrAlaThrThrLeuSerValLeuValIleLeuLeuPheValGlyGly 85  
 Qy 1640 TCTGCTCTTTAATTTGCAATTTATATGACCATAGGATTTCTTCTAGGAACTTTATCGTCT 1699  
 Db 86 SerIlePheAsnPheAlaPheIleMetThrValGlyIleLeuLeuGlyThrLeuSerSer 105  
 Qy 1700 CTTTATATTCACCACCTCTGCTGCTTTGTTATGCTCGCTAAAGAAATCGCTCA 1753  
 Db 106 LeuTyrIleAlaProLeuLeuLeuPheMetValArgLysGluGlnAsn 123  
 RESULT 11  
 AAW98248  
 ID AAW98248 standard; Protein; 525 AA.  
 XX  
 AC AAW98248;  
 XX  
 DT 31-MAR-1999 (first entry)  
 XX  
 DE H. pylori GHPO 1127 protein.  
 XX  
 KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;  
 KW peptic ulcer disease.  
 XX  
 OS Helicobacter pylori.  
 XX  
 FN WO9843478-A1.  
 XX  
 PD 08-OCT-1998.  
 XX  
 PF 01-APR-1998; 98WO-US06371.  
 XX  
 PR 29-JUL-1997; 97US-0902615.  
 PR 01-APR-1997; 97US-0833457.  
 PR 24-JUN-1997; 97US-0881227.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 XX  
 PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;  
 XX  
 DR WPI; 1998-542293/46.  
 XX  
 DR N-PSDB; AAX13967.  
 XX  
 PT New isolated Helicobacter polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of Helicobacter  
 PT infections and gastrointestinal diseases  
 XX  
 PS Claim 8; Page 223-225; 2054pp; English.  
 XX  
 CC This sequence represents a Helicobacter pylori GHPO protein of the  
 CC invention. The polypeptides can be used for preventing or treating  
 CC Helicobacter infections, and gastroduodenal diseases associated with  
 CC these infections, including acute, chronic, and atrophic gastritis, and  
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
 CC used for the production of antibodies. The products can also be used for  
 CC detection and diagnosis.



Db 314 PheSerValAlaGlnAlaSerAspLeuAlaIleAlaLeuArgSerGlyAlaMetAsnAla 333  
Qy 221 GTTCCGAGTTCTCAGTGAAGAGCATCTCTCTGATCTTGGGAAAAACAATGTACA 280  
Db 334 ProIleGlnValLeuGluYbArgIleValGlyProSerLeuGlyLeuAspSerIleLeys 353  
Qy 281 CAAGGCATATCTCAGCATGCTGTGGCTTGGCAATCTTATTTGATGAGCGTATAT 340  
Db 354 ThrSerIleIleAlaLeuValGlyGlyPheIleLeuValMetGlyPheMetAlaLeuYr 373  
Qy 341 TATAGATTGGAGGGCTCATGCTCGGGAGCTGTCTTCTGAATCTTTTGTATCTGG 400  
Db 374 TyrSerMetAlaGlyValIleAlaCysMetAlaLeuValAlaSerLeuPheLeuIleVal 393  
Qy 401 CGAGCTCTACAGTATTTGGATCGCCACTCACCTTGTGAGCACTCGCTGGGATTTCTT 460  
Db 394 AlaValMetAlaIlePheGlyAlaThrLeuThrLeuProGlyMetAlaGlyIleValLeu 413  
Qy 461 GCTATGGGATGGCCGTAGATGCAATGTTCTGTATTTCGAAGATCGGAGGAATTT 520  
Db 414 ThrValGlyIleAlaValAspAlaAsnIleIleIleAsnGluArgIleArgGluValLeu 433  
Qy 521 TTATTGTCTCAAGTCTTAAAGTCTGAGAAAGGATATACCAAGGCTTTTGGAGCC 580  
Db 434 ArgGluGlyGluGlyValValysAlaIleHisLeuGlyTyrIleAsnAlaSerArgAla 453  
Qy 581 ATTTTGTATTCTAAGTCTGACTACAGTATTGGCTCAGCACTTCTTTCTCTAGATACA 640  
Db 454 IlePheAspSerAsnIleThrSerLeuIleAlaSerValLeuLeuYrAlaTyrGlyThr 473  
Qy 641 GGGCTATTAAAGGTTGCTTGTGATGATTTAGGAATTTCTTCAATGTTTACG 700  
Db 474 GlyAlaIleLysGlyPheAlaLeuThrThrGlyIleGlyIleLeuAlaSerIleIleThr 493  
Qy 701 GCTCTTTTCATGACTAAATTTTCTTCATGCTGTGATGAATAAGCAACCAATACACAG 760  
Db 494 AlaIleIleGlyThrGlnGlyIleTyrGlnAlaLeuLeuProLysLeuAlaGlnThrLys 513  
RESULT 13  
AAB88551 standard; Protein; 616 AA.  
XX AAB88551;  
AC AAB88551;  
XX 04-JUN-2001 (first entry)  
XX Haemophilus influenzae essential bacterial protein SEQ ID NO:120.  
XX Haemophilus influenzae; essential bacterial gene; identification;  
KW otitis media; meningitis; upper respiratory tract infection;  
KW infection; antimicrobial.  
XX Haemophilus influenzae.  
XX WO200111033-A2.  
XX 15-FEB-2001.  
XX 03-AUG-2000; 2000WO-US21176.  
XX 04-AUG-1999; 99US-0368382.  
XX (ABBO ) ABBOTT LAB.  
XX Chovan LE, Hessler PE, Reich KA;  
XX WPI; 2001-147511/15.  
DR N-ESDB; AAF94404.  
XX Essential bacterial genes from Haemophilus influenzae and methods for  
PT identifying 'essential' genes that may be potential therapeutic targets  
XX

PS Claim 9; Page 172-173; 185pp; English.  
XX AAF94345 to AAF94409 represent essential bacterial genes from  
CC Haemophilus influenzae, which encode the proteins given in AAB88492 to  
CC AAB88556. The present invention also describes methods for identifying  
CC essential bacterial genes (i.e. those essential to the survival of a  
CC bacterium) using a transposition system. The methods are used to  
CC identify essential genes from bacteria, especially H. influenzae (which  
CC causes otitis media, meningitis and upper respiratory tract infections)  
CC which may be used as targets for potential antimicrobial agents.  
CC AAF94410 to AAF94416 represent PCR primers used in the exemplification  
CC of the present invention.  
XX SQ Sequence 616 AA;  
Alignment Scores: 9.62e-37 Length: 616  
Pred. No.: 408.50 Matches: 87  
Score: 65.85% Conservative: 48  
Percent Similarity: 42.44% Mismatches: 59  
Best Local Similarity: 12.33% Indels: 11  
Query Match: 22 Gaps: 2  
DB:  
US-09-868-987-1 (1-1864) x AAB88551 (1-616)  
Qy 137 AATCATGCGAGTGTCTCAGGGAATTT-----ACC 166  
Db 395 AsnValAlaThrIleGlnGlyArgPheGlySerAsnPheGlnIleThrGlyValAspSer 414  
Qy 167 CACGCTGAAGTGACCAACTCGCTCAGATTAAATCTGGAGCGATGCTTTTGTTCCTC 226  
Db 415 IleAlaGluAlaHisAsnLeuSerThrLeuLysSerGlyAlaLeuIleAlaProIle 434  
Qy 227 GAGGTCTCAGTGAAGAGAGCATCTCTTCGATCTCGGAAACAAATGTACACAAGGC 286  
Db 435 GlnIleValGluGluArgThrIleGlyProSerLeuGlyAlaGlnAsnValGluGlnGly 454  
Qy 287 ATTATCTCATGATCTGTGGCTTGGCAATGCTTATTGTTTTCGAGCGTATATTATAGA 346  
Db 455 IleAsnAlaSerLeuTrpGlyLeuValAlaValIleAlaPheMetLeuPheTyrTyrLys 474  
Qy 347 TTTGGAGCGTATCGCTTCGGAGCTGTTCTCTGTAATCTTTTGTCTTATCTGGCAGCT 406  
Db 475 MetPheGlyValIleAlaSerPheAlaLeuValIleAsnIleValLeuValGlyLeu 494  
Qy 407 CTACAGTATTG---GATGGCCACTCACCTTGTGAGCACTCGCTGGGATTTCTTCTGCT 463  
Db 495 MetSerIleLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeuThr 514  
Qy 464 ATGGGATGGCCGTAGATGCAATGTTCTTGTATTTCGAAAGAAATCCGAGAGGAATTTTA 523  
Db 515 LeuGlyMetSerValAspAlaAsnValIlePheGluArgIleLysGluLeuLeuArg 534  
Qy 524 TTGCTCAAGTCTTAAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCCATT 583  
Db 535 AsnGlyArgSerIleGlnAlaIleAsnGluGlyTyrAsnGlyAlaPheThrSerIle 554  
Qy 584 TTTGATTTCTAAGTCTGACTACAGTATTCGCTCAGCACTTCTTTCTCTAGATACAGGG 643  
Db 555 PheAspAlaAsnLeuThrIleLeuThrAlaIleIleLeuTyrAlaValGlyThrGly 574  
Qy 644 CCTATTAAAGGGTTTGTCTTGCATTTAGGAATTTTCTCTCAATCTTTTACGCT 703  
Db 575 ProIleGlnGlyPheAlaIleThrLeuSerLeuGlyValAlaIleSerMetPheThrAla 594  
Qy 704 CTTTTCATGACTAAA 718  
Db 595 IleThrGlyThrArg 599  
RESULT 14  
AAU91482  
ID AAU91482 standard; Protein; 616 AA.  
XX



AC AAU91482;  
 XX  
 XX 18-JUN-2002 (first entry).  
 XX  
 DE Haemophilus influenzae essential gene #60.  
 XX  
 KM Essential bacterial gene; antifungal agent; antibacterial agent;  
 KM antiparasitic agent; insecticidal agent; microbial infection;  
 KM mucous membrane infection; otitis media; sinusitis; bronchitis;  
 KM alveolitis; conjunctivitis; pneumonia; meningitis; epiglottitis;  
 XX cellulitis; septic arthritis.  
 OS Haemophilus influenzae.  
 XX  
 PN WO200218601-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 22-AUG-2001; 2001WO-US26245.  
 XX  
 PR 25-AUG-2000; 2000US-0649145.  
 XX  
 PA (ABBO ) ABBOTT LAB.  
 XX  
 PI Chovan LE, Hessler PE, Reich KA;  
 XX  
 DR MPI; 2002-304258/34.  
 DR N-PSDB; ABK64968.  
 XX  
 PT Essential bacterial genes in Haemophilus influenzae necessary for  
 PT bacterium's growth and survival, useful for screening inhibitors of  
 PT polypeptides and developing therapeutic agents e.g. antimicrobial  
 XX  
 PS Claim 9; Page 172-173; 185pp; English.  
 XX  
 CC The invention describes an essential bacterial gene (I) comprising a  
 CC purified polynucleotide isolated from Haemophilus influenzae where (I)  
 CC is essential to H. influenzae survival. The encoded polypeptide (II) is  
 CC useful for screening substances that function to inhibit essential H.  
 CC influenzae polypeptides by contacting (II) with the desired substances  
 CC and measuring the response by a screen from specific, enzyme, general,  
 CC affinity, phenotypic and binding screen. (I) and (II) are useful in  
 CC developing therapeutic agents such as antifungal, antibacterial and  
 CC antiparasitic agent, insecticidal agent, and preventive antimicrobial  
 CC agents which are effective in preventing microbial infection or useful  
 CC in treatment of that particular infection. (I) and (II) may also be  
 CC useful in treatment of mucous membrane infections such as otitis media,  
 CC sinusitis, bronchitis, alveolitis, conjunctivitis, pneumonia, meningitis,  
 CC epiglottitis, cellulitis and septic arthritis. This is the amino acid  
 CC sequence of an essential H. influenzae gene, described in the invention.  
 XX  
 SQ Sequence 616 AA;  
 Alignment Scores:  
 Pred. No.: 9.62e-37 Length: 616  
 Score: 408.50 Matches: 87  
 Percent Similarity: 65.85% Conservative: 48  
 Best Local Similarity: 42.44% Mismatches: 59  
 Query Match: 12.33% Indels: 11  
 DB: 23 Gaps: 2  
 US-09-868-987-1 (1-1864) x AAU91482 (1-616)  
 QY 137 AATCATGCCAGTGTCTCAGAGAAATTT-----ACC 166  
 DB 395 AsnValAlaThrIleGlnGlyArgPheGlySerAsnHegInIleThrGlyValAspSer 414  
 QY 167 CACCGTAGAGCAAACTGCGCTCAGATTAAATCTGGAGCGATGCTTTGTTC 226  
 DB 415 IleAlaGluAlaHisAsnLeuSerThrLeuLeuYsSerGlyAlaLeuIleAlaProIle 434  
 QY 227 GAGGTTCAGTAGAAGACGATCTCTTCGATCTTGGAAGAAAACATGACACAGGC 286  
 DB 227 GAGGTTCAGTAGAAGACGATCTCTTCGATCTTGGAAGAAAACATGACACAGGC 286

DB 435 GlnIleValGluGluArgThrIleGlyProSerLeuGlyAlaGlnAsnValGluGlnGly 454  
 QY 287 ATTATCTCAGATGCTGTGGCTTGGCAATGCTATTGTTTGGAGCGCTAATTATPAGA 346  
 DB 455 IleAsnAlaSerLeuThrPrgIleValAlaValIleAlaPheMetLeuPheTyrTyrLys 474  
 QY 347 TTGGAGGGGTGCTATCGCTCCGAGGCTGTTCTTCGAATCTTGTGTTATCTCGGACGT 406  
 DB 475 MetPheGlyValIleAlaSerPheAlaLeuValIleAsnIleValIleLeuValGlyLeu 494  
 QY CTACAGTATTTG--GATCGGCACCTCACCTTGTGCAGACTCGCTGGAGATTGTTTGCT 463  
 DB 495 MetSerIleLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeuThr 514  
 QY 464 ATGGGAGTGGCCGTAGATGCAAAATGTTCTTGTAATTGAAAGAAATCCGAGGAATTTTA 523  
 DB 515 LeuGlyMetSerValAspIleAsnValIleLeuIlePheGluArgIleYsGluGluIleArg 534  
 QY 524 TTGTCGCAAGCTTMAAAATCTGAGAAAAGATATACCAAGGCTTTGGAGCCATT 583  
 DB 535 AsnGlyArgSerIleGlnGlnAlaIleAsnGlyTyrAsnGlyAlaPheThrSerIle 554  
 QY 584 TTGATTTCACTTGACTACAGTATGCGCTCAGACACTTCTTCTCTAGATACAGG 643  
 DB 555 PheAspAlaAsnLeuThrIleThrIleLeuThrAlaIleIleLeuTyrAlaValGlyThrGly 574  
 QY 644 CTTATTAAAGGTTTCTTGGACATTGATTGAAATTTTCTTCAATGTTTACGCT 703  
 DB 575 ProIleGlnGlyPheAlaIleThrLeuSerLeuGlyValAlaIleSerMetPheThrAla 594  
 QY 704 CTTTCATGACTTAA 718  
 DB 595 IleThrGlyThrArg 599  
 RESULT 15  
 AAU91069  
 ID AAU91069 standard; Protein; 618 AA.  
 XX  
 AC AAU91069;  
 XX  
 XX 05-JUN-2002 (first entry)  
 XX  
 DE Neisseria cell surface polypeptide #7.  
 XX  
 KM Cell surface protein; antibacterial; antimicrobial.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO200216612-A2.  
 XX  
 PD 28-FEB-2002.  
 XX  
 PF 21-AUG-2001; 2001WO-GB03759.  
 XX  
 PR 24-AUG-2000; 2000GB-0020952.  
 XX  
 PA (MICK-) MICROSCIENCE LTD.  
 XX  
 PI Lane JD, Hughes MJG, Santangelo JD;  
 XX  
 DR MPI; 2002-280941/32.  
 DR N-PSDB; ABK54085.  
 XX  
 PT Novel peptide encoded by Neisseria meningitidis, useful for manufacture  
 PT of medicament for treatment or prevention of condition associated with  
 PT infection by Neisseria or Gram-negative bacteria -  
 XX  
 PS Claim 5; Page 44-46; 79pp; English.  
 XX  
 CC The invention relates to polypeptides located on the cell surface of  
 CC Neisseria meningitidis, and the polynucleotides encoding them. The  
 CC sequences of the invention are useful for therapeutic or diagnostic use,  
 CC in the manufacture of a medicament for use in treatment or prevention of



CC a condition associated with infection by *Neisseria* or Gram-negative  
CC bacteria. the sequences are also useful for screening potential  
CC antimicrobial drugs or for detection of virulence. Sequences  
CC AAU91063-AAU91079 represent *Neisseria meningitidis* polypeptides of the  
CC invention.

XX  
SQ Sequence 618 AA;

Alignment Scores:	
Pred. No.:	1.08e-31
Score:	364.00
Percent Similarity:	55.95%
Best Local Similarity:	31.75%
Query Match:	10.99%
DB:	23
	Gaps: 3
	Indels: 6
	Mismatches: 105
	Conservative: 61
	Matches: 80
	Length: 618

US-09-868-987-1 (1-1864) x AAU91069 (1-618)

Qy	23	GAGGGATCAGCGCACTGCTATAGCAATAT---TCTGCAACCGTGGATGGCCGTATG	79
Db	348	AspSerAlaGlyGlySerIlePheGlyLeuThrAlaAlaAsnValGlyLyAsArgMet	367
Qy	80	GCTGTAGTGATT-----GACGGTTATATGGTCAGCAGCCCTATTTTAAACGTC	127
Db	368	AlaMetValLeuIleAspGlnGlyLySerSerGluValValThrAlaProValIleArgThr	387
Qy	128	CCATTGAAA---AATCATGCGCAGTGCTCTCAGGGAAATTTACCCACCGTAGTACGCAAA	184
Db	388	AlaIleThrGlyGlyArgValGluIleSerGlySerMetThrThrAlaGluAlaAsnAsp	407
Qy	185	CTGCCTCAGATTTAAATCTGGAGGATGCTCTTTTGTCCCGAGGTTCTCAGTGAAGAG	244
Db	408	ThrSerLeuLeuLeuArgAlaGlySerLeuAlaAlaProMetGlnIleValGluGluArg	427
Qy	245	ACGATCTCTTCGATCTTTGGAAAAACAATGTACACAGGCATTATCTCAGCATGCTGT	304
Db	428	ThrIleGlyProSerLeuGlyLySerGluAsnIleGluLySglyPheHisSerThrLeuTrp	447
Qy	305	GGCTTGGCAATGCTTATGTTTGTATGAGCGGTATATTATAGATTTTGGAGCGTCATCGCT	364
Db	448	GlyPheAlaIleValAlaAlaPheMetValValTyTyArgLeuMetGlyPhePheSer	467
Qy	365	TCGGGACCTGTTCTTCGATCTTTTGTCTTATCTGGGCAGCTCTCAGATTTTGGATGGC	424
Db	468	ThrIleAlaLeuSerAlaAsnIleLeuPheLeuIleGlyLeuSerAlaMetGlnAla	487
Qy	425	CCACTCACCTTGTTCAGACATCGCTGGGATTTCTTCTGATGGGATGCCGTAGATGCA	484
Db	488	ThrLeuThrLeuProGlyMetAlaAlaLeuAlaLeuThrLeuGlyMetAlaIleAspSer	507
Qy	485	AATGTTCTTCTGATTTCGAAAGAAATCCGAGAGGAATTTTATTGTTCTCAAAAGTCTTAAAAA	544
Db	508	AsnValLeuIleAsnGluArgIleArgGluLeuArgAlaGlyValProProGlnGln	527
Qy	545	TCGTAGAAAAAGATATACCAAGCTTTTGGAGCCATTTTGAATCTCACTTGACTACA	604
Db	528	AlaIleAsnLeuGlyPheGlnHisAlaTrpAlaThrIleValAspSerAsnLeuThrSer	547
Qy	605	GTATTGGCCCTCAGCATTCTTTCTTCCTAGATACAGGCCTATTAAAGGTTTGGCTTTG	664
Db	548	LeuIleAlaGlyIleAlaLeuLeuValPheGlySerGlyProValArgGlyPheAlaVal	567
Qy	665	ACATTGATTTTAGGAATTTTCTTCTCAATGTTTACGGCTCTTTTTCAGTCAATATTTTC	724
Db	568	ValHisCysLeuGlyIleLeuThrSerMetTySerSerValValValPheArgAlaLeu	587
Qy	725	TTCATGCTGTGGATGATATAGACCCCAACATACACAG	760
Db	588	ValAsnLeuTrpTyGlyArgArgGtgLyLeuGln	599

RESULT 16

ABG28914

ID ABG28914 standard; Protein; 773 AA.

XX  
AC

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #28905.

XX notes human diagnostic protein #20303.  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX  
OS

XX PN WO200175067-A2.

11-OCT-2001

11-OCT-2001: 2001WO-US08631.  
30-MAR-2001: 2001WO-US08631.

30-MAR-2001; 2001WG-0308031  
XX  
31-MAR-2000; 2000HS-0540217  
PR

PR 23-AUG-2000; 2000US-0649167.  
XX

PA (HYSE-) HYSEQ INC.  
YY

Drmanac RT, Liu C, Tang YT;

XX  
DR WPI; 2001-639362/73.

DR N-PSDB; AAS93101.  
XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX  
PS Claim 20; SEQ ID No 59273; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at <ftp.wipo.int/pub/published> pct sequences.

XX Sequence 773 AA:

Alignment Scores:	3.93e-31	Length:	773
Pred. No.:	359.50	Matches:	80
Score:	62.3%	Conservative:	61
Percent Similarity:	35.4%	Mismatches:	82
Best Local Similarity:	10.85%	Indels:	3
Query Match:	22	Gaps:	2
DB:			

US-09-868-987-1 (1-1864) x ABG28914 (1-773)

44 AATGGACAATATTCTGCAAAACCGTGGATGGCGTATGGCTGTAGTGATTGACGGTTATATG 103

[illegible]

Db 533 SerGlyLysLysAspAlaAsn-----GlyArgAlaValLeuValLysGlnGluVal

```
QY 104 GTCAGCAGCCGCTATTAAAGTCATGTGAAAAATCATGCCAGTGTCTGAGGAATTT 163
Db 551 ILeSnIlLeAlaenIlLeGlnSerArgLeuGlyAsnSerPheAlaGlyIleMGIleAsn 570
QY 164 ACCACCGCTGAAGTACGAAACTGCGCTCAGATTAAATCTGGAGCGAGTCTTTGTT 223
Db 571 AsnProAsnGlnAlaArgGlnLeuSerLeuLeuArgAlaGlyAlaLeuIlLeAlaPro 590
QY 224 CCCGAGGTTCTCAGTGAAGACATCTCTTGTATCTTGGGAAAAACAATGTACACA 283
Db 591 ILeGlnIlLeValGlnGluArgThrIleGlyProThrLeuGlyMetGlnAsnIlLeGln 610
QY 284 GGCATTCTCAGCATGCTGTGGCGCATGCTTATGTTTGAAGAGCGATATATAT 343
Db 611 GlyLeuGlnAlaCysLeuAlaGlyLeuLeuValSerIlLeLeuPheMetIlLeIlePhe 630
QY 344 AGATTGTGAGCGCTCATCGCTTCGAGAGCTGTTCTTGAATCTTTGCTTATCTGGGCA 403
Db 631 LysIlePheGlyLeuIlLeAlaThrSerAlaLeuIlLeAlaAsnLeuIlLeValGly 650
QY 404 GCTCTACAGTATTTG--GATGCCCACTACCTTGTACAGACTCGCTGGATGTTCTT 460
Db 651 ILeMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 670
QY 461 GCTATGGGAGTGGCGGTAGATGCAATGTTCTTGTATTCGAAAGAATCCGAGAGAAAT 520
Db 671 ThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGlnAlaGlyLeuSglnGlnLeu 690
QY 521 TTATTGTCTCAAGCTTTAAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCC 580
Db 691 SerAsnGlyArgThrValGlnGlnAlaIleAspGlnGlyTrpArgGlyAlaPheSerSer 710
QY 581 ATTTTGAATCTTAAGTACAGTACAGTATGGCTCAGACCTCTTCTTCTTCTGAATACA 640
Db 711 ILePheAspAlaAsnIlLeThrThrLeuIlLeValIleIleLeuTrpAlaValGlyThr 730
QY 641 GCGGCTTAAAGGTTTGTGCTTGAATGTGATTTAGGAATTTCTCTCAATGTTTACG 700
Db 731 GlyAlaIleLeuGlyPheAlaIleThrThrGlyIleGlyValAlaIleAsnSerMetPheThr 750
QY 701 GCTCTTTTCAGTAA 718
Db 751 AlaIleValGlyThrArg 756

RESULT 17
ABG26197
ID ABG26197 standard; Protein; 713 AA.
XX
AC ABG26197;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic p:protein #26188.
XX
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI, 2001-639362/73.
```

```
DR N-PSDB; AAS90384.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 56556; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 713 AA;
XX
Alignment Scores:
Pred. No.: 6,34e-31 Length: 713
Score: 357.50 Matches: 75
Percent Similarity: 65.48% Conservative: 54
Best Local Similarity: 38.07% Mismatches: 67
Query Match: 10,79% Indels: 1
DB: Gaps: 1

US-09-868-987-1 (1-1864) x ABG26197 (1-713)
QY 131 TTGAAAAATATCCAGTGTCTCAGGGAATTTACCCACCGTAGAGCAATCTGCC 190
Db 500 LeuGlyAsnSerPheArgIleTrnGlyIleAsnAsnProAsnGlnAlaArgGlnLeuSer 519
QY 191 TCAGATTAAATCTGAGCGATGTCTTTGTTCCGAGGTTCTCAGTGAAGACAGATC 250
Db 520 LeuLeuLeuArgAlaGlyAlaLeuIlLeAlaProIlLeGlnIleValGlnGlnAlaGlyThrIle 539
QY 251 TCTTGTGATCTTGGAAAAACAATGTACACAGCATTTATCTCAGCATGCTGTGCTTG 310
Db 540 GlyProThrLeuGlyMetGlnAsnIlLeGlnGlnGlyLeuGlnAlaCysLeuAlaGlyLeu 559
QY 311 GCATGCTTATTTTGTAGACGTATATATAGATTGGAGCGCATCGCTTGCGGA 370
Db 560 LeuValSerIlLeLeuPheMetIlLeIlePheTrpLysIlePheGlyLeuIlLeAlaTrnSer 579
QY 371 GCTGTTCTTGAATCTTTTGTATCTTGTGAGCGCATCTACAGTATTTG--GATGGCGCA 427
Db 580 AlaLeuIlLeAlaAsnLeuIlLeValGlyIleMetSerLeuLeuProGlyAlaThr 599
QY 428 CTCACCTTGTACGACTCGTGGATTTGTTCTGTATGGGATGCGCGTAGATGCAAT 487
Db 600 LeuSerMetProGlyIleAlaGlyIleValLeuThrLeuAlaValAlaPheAlaAsn 619
QY 488 GTTCTGTATTCGAAGAATCCGAGAGAAATTTTATGTCTCAAGCTCTTAAATAATCT 547
Db 620 ValLeuIleAsnGlnArgIleGlyGlnGlnLeuSerAsnGlyArgThrValGlnGlnAla 639
QY 548 GTAGAAAAGATATACCAAGGCTTTTGGAGCATTTTGTATCTTACGTACAGTACGTA 607
```

Db 640 IleAspGluGlyTyrArgGlyAlaPheSerSerIlePheAspAlaAsnIleThrThrLeu 659  
 Qy 608 TTGGCTCAGCAGCTTTCTTCTCTAGATACAGGCGCTATTAAAGGGTTTGGCTTTGACA 667  
 Db 660 IleLysValIleIleLeuTyrAlaValGlyThrGlyAlaIleLysGlyPheAlaIleThr 679  
 Qy 668 TTGATTTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATGACTAAA 718  
 Db 680 ThrGlyIleGlyValAlaThrSerMetPheThrAlaIleValGlyThrArg 696

RESULT 18  
 AAW98768  
 ID AAW98768 standard; Protein; 323 AA.  
 XX  
 AC AAW98768;  
 XX  
 DT 31-MAR-1999 (first entry)  
 XX  
 DE H. pylori GHPO 1125 protein.  
 XX  
 KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;  
 KW peptic ulcer disease.  
 XX  
 OS Helicobacter pylori.  
 XX  
 FN WO9843478-A1.  
 XX  
 PD 08-OCT-1998.  
 XX  
 PF 01-APR-1998; 98WO-US06371.  
 XX  
 PR 29-JUL-1997; 97US-0902615.  
 PR 01-APR-1997; 97US-0833457.  
 PR 24-JUN-1997; 97US-0881227.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 FA (INWR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 XX  
 FI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;  
 XX  
 DR WPI: 1998-542293/46.  
 DR N-PSDB; AAX14487.  
 XX  
 PT New isolated Helicobacter polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of Helicobacter  
 PT infections and gastrointestinal diseases  
 XX  
 PS Claim 8; Page 1658-1659; 2054pp; English.  
 XX  
 CC This sequence represents a Helicobacter pylori GHPO protein of the  
 CC invention. The polypeptides can be used for preventing or treating  
 CC Helicobacter infections, and gastroduodenal diseases associated with  
 CC these infections, including acute, chronic, and atrophic gastritis, and  
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
 CC used for the production of antibodies. The products can also be used for  
 CC detection and diagnosis.  
 XX  
 SQ Sequence 323 AA;  
 Alignment Scores:  
 Pred. NO.: 2.26e-30 Length: 323  
 Score: 351.00 Matches: 91  
 Percent Similarity: 50.81% Conservative: 66  
 Best Local Similarity: 29.45% Mismatches: 104  
 Query Match: 10.60% Indels: 48  
 DB: 19 Gaps: 7

US-09-868-987-1 (1-1864) x AAW98768 (1-323)

Qy 830 GTTCTCGAAGTGTCTTTCTTTAGTTGGTCTCGGGTTTGGAGCTGGATTC 889  
 Db 22 ValSerAlaIleLeuAlaLeuAlaLeuAlaLeuGlyLeuPhePheLysGlyPheSer--- 40

Qy 890 GTTTTGGGAATGGATTTTAAAGGAGGGTATGCCCTTTACCTTTTAAATCCAAAAGACGATGCC 949  
 Db 41 ---LeuGlyIleAspPheAlaGlyGly----- 48  
 Qy 950 ATCAGCGATGTGCTCAAAATCGCGGCG-----AAAGTTGTGCATAAA 991  
 Db 49 -----SerLeuValGlnValArgTyrThrGlnAsnAlaProIleLysGluValArgAsp 66  
 Qy 992 CTACAGGAAGCTGTCTTTCTTCTAGAGACTTCGGTATTCAACACATTTGGATCTTCAGAA 1051  
 Db 67 LeuPheGluLysGluAlaArgPheLysGlyValGlnValSerGluPheGlySerLysGlu 86  
 Qy 1052 AAGATCAAAATCTATTATTAGTATAAAGCTTTAAGCTATTACTAAGCAGATACGAGCTCT 1111  
 Db 87 GluIleLeuIleLysPheProPheValGluThrAlaGluAsnGluAspLeuAsnAla--- 105  
 Qy 1112 CTCCTAAATTTAAGCATCATGAGCTGCGGTATTGTGGGATTGTGTGCAGAAACAGGCT 1171  
 Db 106 -----IleValAlaAsn----- 109  
 Qy 1172 AGATTTCTCTACGGAAACTCTAAACGAAACGCAAAATTTTGGTCAAAGGTAAAGCAGCAA 1231  
 Db 110 ---IleLeuLysProSerGlyAspPheGluIleArgLysPheAspThrValGlyProArg 128  
 Qy 1232 CTATCGAAGAAATGCGTTATACGCGACCATCGGCTTTTAGGAGCTTTGGCAATCATC 1291  
 Db 129 ValGlySerGluLeuLysGlyLysGlyLeuSerLeuIleLeuAlaIleAlaIle 148  
 Qy 1292 TTGCTCTATGTAGTTTGGCTTTGAATGGCAATATGCTTTCAGTGGCGTATGGCTTTA 1351  
 Db 149 MetValTyrValSerPheArgTyrGluTrpArgPheAlaLeuAlaSerValIleAlaLeu 168  
 Qy 1352 ATTCATGAC-----CTTTTGGCTACCTGTGCGAGTCTTGTATTATAGCACATTTCTTTTG 1405  
 Db 169 ValHisAspValIleLeuValAlaSerSerValIleValPhe----- 182  
 Qy 1406 AAGAAATTCAAATAGATTTCGAAGCCATTGGTCTTTAATGACTGTATGGGGTATTCA 1465  
 Db 183 ---LysIleAspMetAsnLeuGluValIleAlaAlaLeuLeuThrLeuIleGlyTyrSer 201  
 Qy 1466 TTAACATATCTTTGATCATTTTTGGTATTCGTGATGAGATCCCAAGCGAAGCTGTTT 1525  
 Db 202 IleAsnAspThrIleIleIlePheAspArgIleArgGluGluMet\*\*\*SerGlnLysThr 221  
 Qy 1526 ACCCTATGATGTTTGTAGTTAATGATGCCCTTCAAAGACGCTTCAGCGCCACGTAATG 1585  
 Db 222 LysAsnAlaThrGlnAlaIleAspGluAlaIleSerSerThrLeuThrArgThrLeuLeu 241  
 Qy 1586 ACAACAGCTACAACTCTATCAGTTTGTATTATGCTTTTGTATAGCGGCTCTCTCTGTC 1645  
 Db 242 ThrSerLeuThrValPhePheValValIleLeuLeuCysValPheGlySerLysIleIle 261  
 Qy 1646 TTTAATTTTGGATTTATATGACCATAGGATTTCTTCTAGGAAGCTTTATCTCTCTTTAT 1705  
 Db 262 IleGlyPheSerLeuProMetLeuIleGlyThrIleValGlyThrThrSerSerIlePhe 281  
 Qy 1706 ATTGCACCACCTCTGTTGTTGTTTATG 1732  
 Db 282 IleAlaProLysValAlaLeuLeuLeu 290

RESULT 19  
 AAY37782  
 ID AAY37782 standard; Protein; 140 AA.  
 XX  
 AC AAY37782;  
 XX  
 DT 07-OCT-1999 (first entry)  
 XX  
 DE Amino acid sequence of a Chlamydia trachomatis protein.  
 XX  
 KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;

KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.  
 XX Chlamydia trachomatis.  
 OS  
 XX  
 PN W09928475-A2.  
 XX  
 PD 10-JUN-1999.  
 XX  
 XX 27-NOV-1998; 98WO-IB01939.  
 XX  
 XX 04-NOV-1998; 98US-0107077.  
 PR 28-NOV-1997; 97FR-0015041.  
 PR 17-DEC-1997; 97FR-0016034.  
 XX  
 XX (GEST ) GENSET.  
 PA  
 XX  
 PI Griffais R;  
 XX  
 DR WPI; 1999-371125/31.  
 XX  
 XX Genome sequence of Chlamydia trachomatis  
 PT  
 XX  
 PS Disclosure; Page 1368; 1755pp; English.  
 XX  
 CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
 CC of Chlamydia trachomatis (see A201425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conventional trachoma, nonendemic trachoma,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC perihepatitis, bartholinitis; pneumopathy in breast feeding infants;  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.  
 XX  
 XX Sequence 140 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 2,43e-30 Length: 140  
 Score: 349.00 Matches: 82  
 Percent Similarity: 69.12% Conservative: 12  
 Best Local Similarity: 60.29% Mismatches: 32  
 Query Match: 10.92% Indels: 11  
 DB: Gaps: 1  
 US-09-868-987-1 (1-1864) x AAY37782 (1-140)  
 QY 1770 TTTAAGCGTACTTATTTTGAAGCGATT-----TCT 1741  
 Db 5 PheArgIleThrTyrAlaGlnAlaIleTyrPheIleIleTyrArgArgGluPheCysSerSer 24  
 QY 1740 TTTCGGACCATTAACACACAGAGGCGTGCATATATTAAGACGATTAAGTCTCTAGA 1681  
 Db 25 LeuArgThrMetCysAsnLysArgSerGlyGlyAlaMetCylTranGlnLysSerValProAsn 44  
 QY 1680 AGAATCCCTATGTCATATTAATGCAAAATTAAGACAGAGGCGCGCTATTAACAAA 1621  
 Db 45 LysIleProThrValIleMetAsnAlaLysLeuLysIleGluProProPheThrAsnAsn 64  
 QY 1620 AGCATTAACAAACTGATAGAGTTGAGCTGTGTGCTATTACCGTGGCGCTGAAGCTCTT 1561  
 Db 65 LysMetThrSerThrAspArgValAlaValAlaValIleThrValArgProAsnValPhe 84  
 QY 1560 TGAAGGCGATTAATCAATCAATGATAGGGGTAACAGATTGCGTGGCGATCTTCA 1501  
 Db 85 CysAsnLysSerLeuIleLysIleGlyIleGlyValAsnSerPheSerArgArgSerSer 104  
 QY 1500 CGAATACGATCAAAATGATCAAGATATTGTTAATGATATCCCAATATACGATCTTAA 1441  
 Db 105 ArgIleArgSerLysIleIleAsnValLeuPheLysGluTyrProSerThrValIleAsn 124  
 QY 1440 GCACCAATGGCTGCATTAATTTTG-AAATTTCTCAAAAAGAAATG 1394

Db 125 AAlaProIleAlaCysLysSerIleCysAsnSerLeuGlnLysGluMet 140  
 RESULT 20  
 AAB88556  
 ID AAB88556 standard; Protein; 325 AA.  
 XX  
 AC AAB88556;  
 XX  
 XX 04-JUN-2001 (first entry)  
 DT  
 XX  
 XX Haemophilus influenzae essential bacterial protein SEQ ID NO:130.  
 DE  
 XX  
 XX Haemophilus influenzae; essential bacterial gene; identification;  
 KW otitis media; meningitis; upper respiratory tract infection;  
 KW infection; antimicrobial.  
 XX  
 XX Haemophilus influenzae.  
 OS  
 XX  
 XX W0200111033-A2.  
 PN  
 XX  
 PD 15-FEB-2001.  
 XX  
 PF 03-AUG-2000; 2000WO-US21176.  
 XX  
 XX 04-AUG-1999; 99US-0368382.  
 PR  
 XX  
 PA (ABBO ) ABBOTT LAB.  
 XX  
 XX Chovan LE, Heesler PE, Reich KA;  
 PI  
 XX  
 XX WPI; 2001-147511/15.  
 DR N-PSDB; AAF94409.  
 DR  
 XX  
 PT Essential bacterial genes from Haemophilus influenzae and methods for  
 PT identifying 'essential' genes that may be potential therapeutic targets  
 PT  
 XX  
 XX Claim 9; Page 183-184; 185pp; English.  
 PS  
 XX  
 CC AAF94345 to AAF94409 represent essential bacterial genes from  
 CC Haemophilus influenzae, which encode the proteins given in AAB88492 to  
 CC AAB88556. The present invention also describes methods for identifying  
 CC essential bacterial genes (i.e. those essential to the survival of a  
 CC bacterium) using a transposition system. The methods are used to  
 CC identify essential genes from bacteria, especially H. influenzae (which  
 CC causes otitis media, meningitis and upper respiratory tract infections)  
 CC which may be used as targets for potential antimicrobial agents.  
 CC AAF94410 to AAF94416 represent PCR primers used in the exemplification  
 CC of the present invention.  
 CC  
 XX  
 XX Sequence 325 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 2,38e-29 Length: 325  
 Score: 342.00 Matches: 98  
 Percent Similarity: 51.90% Conservative: 80  
 Best Local Similarity: 28.57% Mismatches: 11  
 Query Match: 10.33% Indels: 54  
 DB: Gaps: 15  
 US-09-868-987-1 (1-1864) x AAB88556 (1-325)  
 QY 710 ATGACTAAATTTTCTTCATGCTGTGGATGAATTAAGACCCACATACAGATTGATATG 769  
 Db 1 MetMetLysLeuHe-----ThrLysAspLysAspGlyHisPhe 13  
 QY 770 ATGAATTAAGTTCGTGGGATTAAG-----CATATTTCTTGAAGAGATGC 814  
 Db 14 IleArgGluIleAsnGlyIleLysLeuProPheProLeuThrIleuHeMetLysValArg 33  
 QY 815 AAAAANCTTGGGCGTGTTCGTAAGAGTGTTCCTTTAGCT-----TCGATTGCT 865

Db 34 LysLeuGlyTyrIleLeuSerAlaLeuLeuMetValIleSerLeuPhePheIleIleThr 53  
 Qy 866 CTCGGTTGGAGCCTCGAATCCGTTTGGGAATGGAATTTAAAGAGGTATGCCCTTT 925  
 Db 54 LysGlyPheAsn---Trp-----GlyLeuAspPheThrGlyValValPhe 68  
 Qy 926 ACCTTTAAATCCAAAGACATGGCATCAGCATGTGTCTCAATCGCTGGCAAGTTGTG 985  
 Db 69 AspThrHisPheSerGlnSer-----AlaAsnLeuGluGlnIleArgSer----- 83  
 Qy 986 CATAAATACAGGAGCTGGTCTCTCTAGAGACTTCCTGATTTCAAAACATTTGGATCT 1045  
 Db 84 ---LysLeuHisGluAsnGlyIleGluSerPro-----IleValGlnThrThrGlySer 100  
 Qy 1046 TCAGAAAGATCAAAATCTATTTAGTGATAAAGCTTTTAAGCTATATAAGCAGATACGA 1105  
 Db 101 ValGlnAspValMetIle-----ArgLeuPro 109  
 Qy 1106 GCCTCTCTCTAAATTAACGATCATGAGCTGGCGCTTATTGTGGGATTTGTCTCAAAAC 1165  
 Db 110 AlaSerAsnAsnAspSerThrIle-----GlyGluHisValLysSerMetLeuGlnAsn 127  
 Qy 1166 AGGCTAGATTTCTACGGAACCTCTAAACGAAACGCAAAATTTTGTCAAGGTAAAGC 1225  
 Db 128 ValAspLysAspIle-----GlnIleArgSerIleGluPhe-----ValGly 141  
 Qy 1286 AGCAAACTATCGAAGAAATCGTTTATCAGCGACCATCGGCTTTTAGGAGCTTTGGCA 1285  
 Db 142 ProAsnValGlyGluLeuAlaGlnGlyAlaValTyrAlaThrLeuAlaThrLeuAla 161  
 Qy 1286 ATCATCTGCTCTATGAGTTGGCTTTGATGATGCAATATGCTTTCAGTGCCTATGC 1345  
 Db 162 MetValLeuIleTyrValGlySerArgPheGluTrpArgLeuGlyPheGlySerIleAla 181  
 Qy 1346 GCTTTAATTCATGACCTTTTGCTACCTGTCAGCTGTTGTTTATAGCACATTTCTTTTG 1405  
 Db 182 SerLeuAlaHisAspValIleThrLeuGlyVal-----PheSerAla 196  
 Qy 1406 AAGAAATTCAAATAGATTGCAAGCCATTTGCTTTAATGACTGTTATGGGGTATTCA 1465  
 Db 197 LeuGlnIleGluIleAspLeuThrPheValAlaAlaIleLeuSerValValGlyTyrSer 216  
 Qy 1466 TTAACATATCTTCATCATTTTTCATCGTATTCGTGAAGAT---CGCAAGCGAACCTG 1522  
 Db 217 IleAsnAspSerIleValValPheAspArgValArgGluAsnPheArgLysIleArgArg 236  
 Qy 1523 TTTACCCCTATGCTGTTTGTAGTTAATGATGCTTCAAAAGACCTTCAGCGCAGCGTA 1582  
 Db 237 LeuAspThrIleAspIle---IleAspIleSerLeuThrGlnThrLeuSerArgThrIle 255  
 Qy 1583 ATGCAACAGCTACAACTCTATCAGTTTGTGTTAATGCTTTTGTATAGGCGGCTCCTCT 1642  
 Db 256 IleThrSerValThrThrLeuValValValMetAlaLeuPhePheGlyGlyProSer 275  
 Qy 1643 GTCCTTAATTTGCACTTATTATGACATAGGATTTCTTAGGAACCTTATCGTCTCTT 1702  
 Db 276 IleHisAsnPheSerLeuAlaLeuLeuValGlyIleGlyPheGlyThrTyrSerSerIle 295  
 Qy 1703 TATATTGCA 1711  
 Db 296 PheValAla 298  
 RESULT 21  
 ID AAU91487 standard; Protein; 325 AA.  
 XX  
 AC AAU91487;  
 XX  
 DT 18-JUN-2002 (first entry)  
 XX  
 DE Haemophilus influenzae essential gene #55.  
 XX  
 KW Essential bacterial gene; antifungal agent; antibacterial agent;

KW antiparasitic agent; insecticidal agent; microbial infection;  
 KW mucous membrane infection; otitis media; sinusitis; bronchitis;  
 KW alveolitis; conjunctivitis; pneumonia; meningitis; epiglottitis;  
 XX cellulitis; septic arthritis.  
 OS Haemophilus influenzae.  
 XX  
 PN W0200218601-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 22-AUG-2001; 2001WO-US26245.  
 XX  
 PR 25-AUG-2000; 2000US-0649145.  
 XX  
 PA (ABBO ) ABBOTT LAB.  
 XX  
 PI Chovan LE, Hessler PE, Reich KA;  
 XX  
 DR WPI; 2002-304258/34.  
 DR N-PSDB; ABK64973.  
 XX  
 PT Essential bacterial genes in Haemophilus influenzae necessary for  
 bacterium's growth and survival, useful for screening inhibitors of  
 PT polypeptides and developing therapeutic agents e.g. antimicrobial -  
 XX  
 PS Claim 9; Page 183-184; 185pp; English.  
 XX  
 CC The invention describes an essential bacterial gene (I) comprising a  
 CC purified polynucleotide isolated from Haemophilus influenzae where (I)  
 CC is essential to H. influenzae survival. The encoded polypeptide (II) is  
 CC useful for screening substances that function to inhibit essential H.  
 CC influenzae polypeptides by contacting (II) with the desired substances  
 CC and measuring the response by a screen from specific, enzyme, general,  
 CC affinity, phenotypic and binding screen. (I) and (II) are useful in  
 CC developing therapeutic agents such as antifungal, antibacterial and  
 CC antiparasitic agent, insecticidal agent, and preventive antimicrobial  
 CC agents which are effective in preventing microbial infection or useful  
 CC in treatment of that particular infection. (I) and (II) may also be  
 CC useful in treatment of mucous membrane infections such as otitis media,  
 CC sinusitis, bronchitis, alveolitis, conjunctivitis, pneumonia, meningitis,  
 CC epiglottitis, cellulitis and septic arthritis. This is the amino acid  
 CC sequence of an essential H. influenzae gene, described in the invention.  
 XX  
 SQ Sequence 325 AA;  
 Alignment Scores:  
 Pred. No.: 2,38e-29 Length: 325  
 Score: 342.00 Matches: 98  
 Percent Similarity: 51.90% Conservatve: 80  
 Best Local Similarity: 28.57% Mismatches: 111  
 Query Match: 10.33% Indels: 54  
 DB: 23 Gaps: 15  
 US-09-868-987-1 (1-1864) x AAU91487 (1-325)  
 Qy 710 ATGACTAAATTTTCTTTCATGCTGTCATGAATAAGCAACATACACAGTTGCATATG 769  
 Db 1 MetMetLysLeuPhe-----ThrLysAspLysAspGlyHisPhe 13  
 Qy 770 ATGAATAAGTTCGTGGGATAAAG-----CATGATTTCTTGAGAGGATGC 814  
 Db 14 IleArgGluIleAsnGlyIleLysLeuProPheProLeuThrGluPheMetLysValArg 33  
 Qy 815 AAAAACTTTGGCGCTGTTTCTGGAGTGTCTTTCTTTAGT-----TGGTGTCT 865  
 Db 34 LysLeuGlyTyrIleLeuSerAlaLeuLeuMetValIleSerLeuPhePheIleIleThr 53  
 Qy 866 CTCGGTTTGGAGCCTCGAATTCGCTTTTGGGAATGGAATTTAAAGAGGATGATGCCCTTT 925  
 Db 54 LysGlyPheAsn---Trp-----GlyLeuAspPheThrGlyValValPhe 68  
 Qy 926 ACCTTTAATCCAAAGAGCATGGCATGCTGTCATGCTGTCATGCTGTCATGCTGTCATG 985

Db 69 AsphrhrhspheSerGlnSe-----AlasnhleuGluglnIleArgSer----- 83  
 Qy 986 CATAAACTACAGGAAGCTGGCTCTTCTTCTAGAGACTTCGCTATTCACCAATTGGATCT 1045  
 Db 84 --LysleuHleGlValasnGlyIleGlnSerPro-----IleValGlnThrThylGlySer 100  
 Qy 1046 TCAGAAAAGATCAAAATCTATTTTAGTCATTAAGCTTTACTAGACGATGCA 1105  
 Db 101 ValGlnAspValMetIle-----ArgLeuPro 109  
 Qy 1106 GCGCTCTCCCTAAATTAATACGATCATGAGCTGGCTTATTTGGAGATGGTTGTCAGAAAC 1165  
 Db 110 AlasSerAsnaAsnAspSerThrIle-----GlyGlnHisValIlySerMetLeuGlnAsn 127  
 Qy 1166 AGGCGCTAGATTCTCTACGGAACCTTAACGAAACGCAAAATTTGGTCAGAGTAAGC 1225  
 Db 128 ValAspLysAspIle-----GlnIleArgSerIleGlnPhe-----ValGly 141  
 Qy 1226 AGCAAACTATGAGAAAATGCGTTATTCAGCGGACGACCTGGGCTTTTAGAGCTTTGGCA 1285  
 Db 142 ProAsnValGlyGlnGluLeuAlaGlnGlyAlaValIlyrAlaThrLeuAla 161  
 Qy 1286 ATCATCTGCTCTATGATGAGTTGGCTTTGATAGGCAATATGCTTCAGTCCGCTATGC 1345  
 Db 162 MetValIleuIleIyValGlySerArgPheGlnThrPalaGlnGlyPheGlySerIleAla 181  
 Qy 1346 GCTTTAATTCATGACCTTTTGGCTACCTGTGACGTCTTGTATTATAGACATTTCTTTTG 1405  
 Db 182 SerLeuAlaHleAspValIleIleThrLeuGlyVal-----PheSerAla 196  
 Qy 1406 AAGAAATTCATCAATGATTTGGACACCTGGCTTTAATGACTGTTATGGCTATTC 1465  
 Db 197 LeuGlnIleGlnIleAspLeuThrPheValAlaAlaIleLeuSerValValGlyIlyrSer 216  
 Qy 1466 TTAACCAATCTCTGATCATTTTGTATCGTGAAGT---CGCCAGACGAACCTG 1522  
 Db 217 IleAsnAspSerIleValValPheAspArgValArgGluAsnPheArgGlyIleArgArg 236  
 Qy 1523 TTTACCCCTGATCATGTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1582  
 Db 237 LeuAspThrIleAspIle---IleAspIleSerLeuThrGlnThrLeuSerArgThrIle 255  
 Qy 1583 ATGACACAGCTACACTCTCATCAGTTTGTATGATGCTTTGTATTATAGGGGCTCTCT 1642  
 Db 256 IleThrSerValThrThrLeuValValAlaMetAlaLeuPhePheGlyIlyrProSer 275  
 Qy 1643 GTCCTTAATTTGGCATTTATATATGACCATGAGGATTTCTTAGAGAACTTATCGCTCTT 1702  
 Db 276 IleHlsAsnPheSerLeuAlaLeuLeuValGlyIleGlyPheGlyThrIlyrSerSerIle 295  
 Qy 1703 TATATTGCA 1711  
 Db 296 PheValAla 298  
 RESULT 22  
 ID AU63631 standard; Protein; 576 AA.  
 AC AU63631;  
 DT 27-FEB-2002 (first entry)  
 DE Propionibacterium acnes immunogenic protein #24527.  
 KM SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;  
 KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KM dermatological; osteopathic; neuroprotectant.  
 OS Propionibacterium acnes.  
 PN WO200181581-A2.

PD 01-NOV-2001.  
 XX 20-APR-2001; 2001WO-US12865.  
 PF 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX (CORI-) CORIXA CORP.  
 PA Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 DR MPI: 2001-616774/71.  
 DR N-PsDB; AAS59634.  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 PS Claim 3; SEQ ID No 24826; 1069bp; English.  
 XX Sequences AU93105-AU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 576 AA;  
 Alignment Scores:  
 Pred. No.: 3,45e-25 Length: 576  
 Score: 306.50 Matches: 74  
 Percent Similarity: 56.44% Conservative: 53  
 Best Local Similarity: 32.89% Mismatches: 87  
 Query Match: 9.25% Gaps: 11  
 DB: 22  
 US-09-868-987-1 (1-1864) x AU63631 (1-576)  
 Qy 74 CGTAGCGCTGATGATGACGTTATATGCTACGACGCTTATTTAAAC----- 124  
 Db 297 GlnPheAlaValIleuAspGlyValIleSerSerProGlnIleuAsnGlyAsnThr 316  
 Qy 125 -----GTCCCAATTGAAA---AATCATGCCAGTGTCTCAGGGAATTTACCAACCGCT 172  
 Db 317 GlyThrSerCysProIleasnGlyGlnValaGlnIleSerGlyHisPheThrGlnAsn 336  
 Qy 173 GAAGTAGCAAACTCGCTCAGATTTAAATCTGAGAGCATGTCTTTGTTCGAGGTT 232  
 Db 337 SerAlaAlaAspLeuAlaAsnValLeuIyryrGlyAlaLeuProLeuSerPheAspIle 356  
 Qy 233 CTCAGTGAAGAGACATCTCTTCTGATCTTTGGGAAAACAATGTATACAAAGCATATTC 292  
 Db 357 SerSerValAspAsnIleSerProThrLeuGlyGlnGlnIleuArgAlaGlyIleIle 376  
 Qy 293 TCAGCATGCTGTGGCTTGCAATGCTTATTTGTTGATGACGATATATTAAGATTGGA 352

```

Db 377 AlaGlyIleIleGlyLeuValGlyGlyTyrCysLeuLeuTyrTyrArgGlyLeu 396
Qy 353 GCGCTCATCGCTCGGGAGCTGTTCTTCTGAATCTTTGCTTATCTGGCAGCTCTACAG 412
Db 397 GlyLeuValThrLeuGlySerLeuValIleAlaGlyIleThrThrTyrAlaAlaMetVal 416
Qy 413 TATTGGATGG-----CCATCCTTGTGCTGAGCTCGCTGGGATGTTCTT 460
Db 417 LeuLeuGlyGluAlaValGlyPheThrLeuSerLeuAlaGlyIleAlaGlyAlaIleVal 436
Qy 461 GCTATGGGATGGCGGTAGTCAAAATCTTCTGTATTCGAAAGAAATCCGAGAGGAATT 520
Db 437 AlaIleGlyValThrAlaAspSerPheIleValTyrPheGluArgIleArgAspGluIle 456
Qy 521 TTATTGTCTCAAGTCTTAAATAATCTGTAGAAAGATATACCAAGCTTTTGGAGCC 580
Db 457 ArgGluGlyArgThrLeuArgThrAlaLeuGlnThrGlyTyrAlaLysAlaGlyThr 476
Qy 581 ATTTTGTATTCTAATCTGACTACAGTATTGGCTGAGCTCTTCTTCTTCTCTAGATACA 640
Db 477 IleLeuMetAlaAspGlyValSerLeuLeuSerAlaIleValLeuPheValLeuSerVal 496
Qy 641 GGGCTATTAAAGGTTTGTCTTGCATTGATTTTGGAAATTTTCTCTCAATGTTTACG 700
Db 497 AspGlnValGlyPheAlaPheThrLeuGlyLeuThrValIleAspLeuIleIle 516
Qy 701 GCTCTTTTCATGACT 715
Db 517 CysPhePheThr 521

RESULT 23
ID AAU91065 standard; Protein, 311 AA.
XX AC AAU91065;
XX DT 05-JUN-2002 (first entry)
XX DE Neisseria cell surface polypeptide #3.
XX KW Cell surface protein; antibacterial; antimicrobial.
XX OS Neisseria meningitidis.
XX PN WO200216612-A2.
XX PD 28-FEB-2002.
XX PF 21-AUG-2001; 2001WO-GB03759.
XX PR 24-AUG-2000; 2000GB-0020952.
XX PA (MICR-) MICROSCIENCE LTD.
XX PI Lane JD, Hughes MJG, Santangelo JD;
XX DR WPI; 2002-280941/32.
XX DR N-PSDB; ABK54081.
XX PT Novel peptide encoded by Neisseria meningitidis, useful for manufacture
XX PT of medicament for treatment or prevention of condition associated with
XX PS infection by Neisseria or Gram-negative bacteria -
XX PS Claim 5; Page 22-24; 79pp; English.
XX CC The invention relates to polypeptides located on the cell surface of
XX CC Neisseria meningitidis, and the polynucleotides encoding them. The
XX CC sequences of the invention are useful for therapeutic or diagnostic use,
XX CC in the manufacture of a medicament for use in treatment or prevention of
XX CC a condition associated with infection by Neisseria or Gram-negative
XX CC bacteria. The sequences are also useful for screening potential
XX CC antimicrobial drugs or for detection of virulence. Sequences

```

```

CC AAU91063-AAU91079 represent Neisseria meningitidis polypeptides of the
XX invention.
SQ Sequence 311 AA;
Alignment Scores: 2,96e-24 Length: 311
Score: 297.00 Matches: 81
Percent Similarity: 52.41% Conservative: 82
Best Local Similarity: 26.05% Mismatches: 104
Query Match: 8.97% Indels: 44
DB: 23 Gaps: 10
US-09-868-987-1 (1-1864) x AAU91065 (1-311)
Qy 839 AGTGTCTTTCTTTTGGTTCGTTGCTCTCGGGTTTGGAGCTGGAATCCGTTTGGGA 898
Db 30 AlaValPhePheLeuValThrArgGlyLeuAsnPhe-----Ser 42
Qy 899 ATGGAATTTAAAGAGGGTATGCTTTTACCTTTAAATCCAAAGAGCATGGCATCAGCAT 958
Db 43 ValGluPheThrGlyGly-----ThrValMetGlu 52
Qy 959 GTTGCTCAAAATGCGTGCACAAAGTTGTGCATAAACTACAGGAAGCT-----GGTCTTTCT 1012
Db 53 ValGlnTyrGlnGlnGlyAlaAspValAsnLysMetArgGluArgLeuAspThrLeuLys 72
Qy 1013 TCTAGAGACTCCGTATTCAAAACATTTGGATCTTCACAAAAGATCAAAATCTATTTAGT 1072
Db 73 IleGlyAspValGlnValGlnAlaLeuGlyThrAsnLysHisIleMetIleArgLeuPro 92
Qy 1073 GATAAAGCTTTAAGCTATATAAGCAGATAGAGCTCTCTCTCTAAATTAACGATCATG 1132
Db 93 AsnLys-----GluGlyValThrSerAlaGlnLeuSerAsnGlnValMet 107
Qy 1133 AGCTGGGCTTATTGTGGGATTTGTCTCAGAAACAGGCTAGATTCTCTACGGAACCTCT 1192
Db 108 Asp-----LeuLeuLysLysAspSerProAspVal-----Thr 118
Qy 1193 AAACGAAACGCAAAATTTTGGTCAAAAGTAAAGCAGCAGCAAACTATCGAAGAAATCGCTTAT 1252
Db 119 LeuArgGlnValGluPhe-----IleGlyProGlnValGlyGluLeuValSer 135
Qy 1253 CAGCGCAGCATCGGGCTTTTAGGAGCTTTGGCAATCATCTTGCTCTATGTAGTTTGGCG 1312
Db 136 AsnGlyLeuMetAlaLeuGlyPheValIleGlyIleIleIleTyrLeuSerMetArg 155
Qy 1313 TTGAATGGCAATATGCTTTTCAGTGGCGTATGCGCTTAAATTCATGACCTTTTGGCTACC 1372
Db 156 PheGluTrpArgPheAlaValSerAlaIleIleAlaAsnMetHisAspIle----- 172
Qy 1373 TGTGCAGTCTTGTATTATAGCACATTTCTTTTGAAGAAATTTCAATAGATTTCGCAAGCC 1432
Db 173 -----ValIleIleLeuGlyCysPheAlaPhePheGlnTrpGluPheSerLeuThrVal 190
Qy 1433 ATTGCTGCTTTAATGACTGTATTGGGGTATTTCATTAACAATACTTTTGATCATTTTGTAT 1492
Db 191 LeuAlaGlyIleLeuAlaValLeuGlyTyrSerValAsnGluSerValValPheAsp 210
Qy 1493 CGTATTCGTGAAGAT-----CGCCAAAGCGAACCTGTTTACCCCTATGCTGTTTGTAGTT 1546
Db 211 ArgIleArgGluAsnPheArgLysProAlaMetArgGlyHisAlaValProGluValIle 230
Qy 1547 AATGATGCCCTTCAAAAGAGCTTCAGCGCGCATGTAATGACACAGCTACAACTCTATCA 1606
Db 231 AspAsnAlaIleThrAlaThrMetSerArgThrIleIleThrHisGlySerThrGluAla 250
Qy 1607 GTTTTGTAAATGCTTTTGTATTATAGGGGCTCTCTCTCTCTTAAATTTTGCATTTATTATG 1666
Db 251 MetValValSerMetLeuValPheGlyAlaAlaLeuHisGlyPheSerMetAlaLeu 270
Qy 1667 ACCATAGGATTTCTTAGGAACCTTTATCGTCTCTTTATATTGACCACCTCTCTGTTGT 1726

```







Db 525 ValPheAspLeuValValThrPheLeu 533

RESULT 25

AAE15862

ID AAE15862 standard; Protein; 637 AA.

XX AC AAE15862;

XX 26-MAR-2002 (first entry)

DT

XX Corynebacterium glutamicum secD protein.

DE

XX Genetically modified bacterial strain; secD, secF, reporter system;

KW enhanced secretion activity.

XX

XX Corynebacterium glutamicum.

OS

XX Key Location/Qualifiers

FT Misc-difference 1 /note= "Encoded by TTG"

XX

XX WO200185967-A2.

XX

XX 15-NOV-2001.

XX

XX 26-APR-2001; 2001WO-EP04703.

XX

XX 12-MAY-2000; 2000EP-0110021.

XX

XX (DEGS ) DEGUSSA AG.

XX

XX Berens S, Kalinowski J, Puehler A;

XX

XX WPI; 2002-082901/11.

DR

XX N-PSDB; AAD25590.

XX

XX Genetically modified Corynebacterium with enhanced secretion activity

PT useful for production of desired substance e.g. protein, comprises a

PT modification at one of the genes secD and secF -

XX

XX Claim 9; Page 38-40; 42pp; English.

XX

XX The present invention relates to genetically modified bacterial strain

CC Corynebacterium glutamicum, comprising a genetical modification at one of

CC the genes secD and secF. The genetically modified bacterial strain is

CC useful for production of desired substance which is an amino acid.

CC oligopeptide, polypeptide or protein preferably a heterologous protein,

CC where the produced substance is secreted by the bacterial strain. The

CC invention is useful in a reporter system. Modification in secD and secF

CC in genetically modified bacterial strain Corynebacterium glutamicum,

CC results in enhanced secretion of the strain, which is utilised for

CC production of high amounts of desired substances which can be easily

CC isolated from the source of production. The present sequence is

CC Corynebacterium glutamicum secD protein.

XX

SO Sequence 637 AA;

Alignment Scores:

Pred No.:	7.08e-22	Length:	637
Score:	277.50	Matches:	67
Percent Similarity:	51.81%	Conservative:	62
Best Local Similarity:	26.91%	Mismatches:	87
Query Match:	8.38%	Indels:	33
DB:	23	Gaps:	7

US-09-868-987-1 (1-1864) x AAE15862 (1-637)

Qy 5 ACTTCCGCATATGTTCAGGAGGGGATGACGGGCACTGCTAATGGACAATATTTCTGCAAAAC 64

||||| ||| |||

Db 326 ThrSerGlnTyrLeuGlnGln----- 332

Qy 65 CGTGGATGGCTATGCTGTAGTGTACGGGTATATGTCAGCAGCCCTATTTTAAAC 124

||||| ||| |||

||||| ||| |||

Db 333 -----GlnIleAlaIleThrLeuAspSerGlnValIleSerAlaProValIleGln 349

Qy 125 -----GTCCCATTTGAAATCATGCCAGTGTCTCAGGGAAATTTACCCACCGTGAAGTG 178

||||| ||| |||

Db 350 SerAlaThrProValGlySerAlaThrSerIleThrGlyAspPheThrGlnThrGluAla 369

||||| ||| |||

Qy 179 AGCAAACTCGCTCAGATTTAAATCTGGAGCG-----ATGCTCTTTTGTCCCGAG--- 229

||||| ||| |||

Db 370 GlnAspLeuAlaAsnAenLeuArgTyrGlyAlaLeuProLeuSerPheAlaGlyGluAsn 389

||||| ||| |||

Qy 230 -----GTTCTCAGTGAAGACGATCTCTCTCATCTTGGGAAAAACAATGTACA 280

||||| ||| |||

Db 390 GlyGluArgGlyThrThrThrValProProSerLeuGlyAlaAlaSerLeuLys 409

Qy 281 CAAGGCATTTATCTCAGCATGCTGTGGCTTGCCATGCTTATTTTTCATGAGCGTATAT 340

||||| ||| |||

Db 410 AlaGlyLeuIleAlaGlyIleValGlyIleAlaLeuValAlaIlePheValPheAlaTyr 429

||||| ||| |||

Qy 341 TATAGATTGGAGCGCTATC-----GCTTCGGGAGCTGTTCTTCTTG 382

||||| ||| |||

Db 430 TyrArgValPheGlyPheValSerLeuPheThrLeuPheAlaAlaGlyValLeuValTyr 449

Qy 383 AATCTTTTGTCTTATCTGGCAGCTCTACAGTATTTGGATGCGCCACTCACCTTCTCAGGA 442

||||| ||| |||

Db 450 GlyLeuLeuValLeuLeuGly-----ArgTrpIleGlyTyrSerLeuAspLeuAlaGly 467

Qy 443 CTCGCTGGGATTGTTCTTGTCTATGGGATGCCGCTAGATGCAAAATGTTCTTGTATCGAA 502

||||| ||| |||

Db 468 IleAlaGlyLeuIleIleGlyIleGlyThrAlaAspSerPheValValPheTyrGlu 487

Qy 503 AGAATCCGAGAGGAATTTTATTTCTCTCAAGTCTTAAATAATCTGTAGAAAAAGGATAT 562

||||| ||| |||

Db 488 ArgIleLysAspGluIleArgGluGlyArgSerPheArgSerAlaValProArgAlaTrp 507

Qy 563 ACCAAGGCTTTTGAGCCATTTTGTATTCTAACTTGAATGACTACAGTATTGGCCTCAGCACTT 622

||||| ||| |||

Db 508 GluSerAlaLysArgThrIleValThrGlyAsnMetValThrLeuLeuGlyAlaIleVal 527

Qy 623 CTTTTCCTCTCTAGATACAGGCGCTTATTAAGGGTTTCTTTGACATTCG---ATTTTAGGA 679

||||| ||| |||

Db 528 IleTyrLeuLeuAlaValGlyGluValLysGlyPheAlaPheThrLeuGlyLeuThrThr 547

Qy 680 ATTTTCCTCTCAATGTTTACGGCTCTT 706

||||| ||| |||

Db 548 ValPheAspLeuValValThrPheLeu 556

AC AAG91576;

XX

XX 26-SEP-2001 (first entry)

XX

XX C glutamicum protein fragment SEQ ID NO: 5330.

DE

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis.

XX

XX Corynebacterium glutamicum.

OS

XX EPI108790-A2.

XX

XX 20-JUN-2001.

XX

XX 18-DEC-2000; 2000EP-0127688.

XX

XX 16-DEC-1999; 99JP-0377484.

PR

XX 07-APR-2000; 2000JP-0159162.

PR

XX 03-AUG-2000; 2000JP-0280988.

XX

XX (KYOW ) KYOWA HAKKO KOGYO KK.

PA

XX

PI	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX	WPI, 2001-376931/40.
DR	N-PSDB; AAH66795.
XX	
PT	Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT	mutation point of a gene, measuring expression of a gene, analysing
PT	expression profile or pattern of a gene and identifying homologous gene
XX	
PS	Claim 17, SEQ ID NO: 5330; 246pp + Sequence Listing; English.
XX	
CC	The present invention provides a number of nucleotide and protein
CC	sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC	are useful for identifying the mutation point of a gene derived from a
CC	mutant of corynebacterium bacterium, measuring expression amount and
CC	analysing the expression profile or expression pattern of a gene derived
CC	from Corynebacterium bacterium, and identifying a homologue of a gene derived
CC	from Corynebacterium bacterium. Corynebacterium bacteria are useful for producing
CC	amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC	particularly L-lysine. The present sequence is a protein described
CC	in the exemplification of the invention.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from the
CC	European Patent Office.
XX	
Q0	Sequence 644 AA;

Alignment Scores:	
Pred. No.:	7 13e-22
Score:	277.50
Percent Similarity:	51.61%
Best Local Similarity:	26.81%
Query Match:	8.38%
DB:	22
	Gaps: 7
	length: 644
	Matches: 67
	Conservative: 62
	Mismatches: 87
	Indels: 33
	Gaps: 7

US-09-868-987-1 (1-1864) X AAC91576 (1-644)

[illegible]

Db 475 IleAlaIleuIleIleIleGlyIleGlyThrTrnAlaAepSerPheValValPheTyrGlu 494

Qy 503 AGAATCCGAGAGATATTTTATTTATGTCTCAAACTCTTAAATAATCTGTGAGAAAAAGCAT 562

Db 495 ArgIleIysAspGluIleArgGluGlyArgSerPheArgSerAlaValProArgAlaTrp 514

Qy 563 ACCGAGGCTTTTGAGGCATTTTGGATTTCATTACCTTGACTACAGTATTCGCTCAGCACTT 622

Db 515 GluSerIleAlaYsArgThrIleValTrnGlyAamMetValTrnLeuLeuGlyAlaIleVal 534

Qy 623 CTTTTCCTTCCTAGATACAGGCGCTATTAAAGGCTTTCCTTGACACTTG--ATTTTACGA 679

Db 535 IleTyrLeuLeuAlaValGlyGluValIysGlyPheAlaPheThrLeuGlyLeuThrThr 554

Qy 680 ATTTTCTCTCATATTTTACGCTCTT 706

Db 555 ValPheAspLeuValValTrnPheLeu 563

RESULT 27

ID AAB05947 standard; Protein; 279 AA.

AA05947

AC AAB05947;

AD 20-OCT-2000 (first entry)

DE Protein deduced from Mycoplasma hyopneumoniae genomic clone pAB727.

KM Mycoplasma hyopneumoniae; vaccine; antigen; antimicrobial;

KW enzootic pneumonia.

OS Mycoplasma hyopneumoniae.

PN WO200031115-A1.

PD 02-JUN-2000.

PF 19-NOV-1999; 99WC-AU01035.

PR 20-NOV-1998; 98AU-0007273.

PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.

PI Moore RJ, Doran TJ;

DR WPI: 2000-400031/34.

DR N-PSDB; AAA56833.

PT Identifying antigenic polypeptides for use in vaccines against

PT Mycoplasma infection comprises screening proteins co-expressed with a

PT marker -

PS Claim 27; Fig 14; 241pp; English.

XX

XX The present sequence is encoded by a clone selected from a Mycoplasma

XX hyopneumoniae genomic DNA library. M. hyopneumoniae causes Enzootic

XX pneumonia in pigs. The infection rarely causes death but often results in

XX significant depression, leading to reduced weight gain. The animals are

XX also prone to secondary infection by opportunistic pathogens. Genomic

XX DNA from M. hyopneumoniae was partially digested with restriction

XX enzymes and fragments were ligated downstream of the polyHis tag region

XX in the expression vector pCI. The ligation mixture was used to transform

XX competent E. coli cells and polyHis positive clones were selected using

XX anti-polyHis antibodies. Expression of the recombinant vector

XX generates polyHis fusion proteins which are easily identified and

XX isolated. Pigs were vaccinated with plasmid DNA from polyHis

XX positive clones in order to identify clones encoding antigenic proteins

XX which confer protection against M. hyopneumoniae infection. Such

XX proteins may subsequently be used in the manufacture of effective

XX vaccines.

Sequence 279 AA;

Alignment Scores:

Pred. No.:	1,94e-18	Length:	279
Score:	245.50	Matches:	77
Percent Similarity:	49.50%	Conservative:	73
Best Local Similarity:	25.41%	Mismatches:	105
Query Match:	7.41%	Indels:	49
DB:	21	Gaps:	9

US-09-868-987-1 (1-1864) x AAB05947 (1-279)

Qy	893	TTGGGAATGATTTTAAAGAGGGTATGCTTTTAACTTAACTCAAAAGAGCATGGCATC	952
Db	8	LeuAlaileAaspPheSerGlyThrAenPheLeuIleGlu-SerSerAenSerSerty	27
Qy	953	AGCGATCTTGCTCAAAATCGTCGCAAAAGTTGTGCATAAATACACAGGAAGCTGGCTTTCT	1012
Db	27	rAspLeuIleThrLysGluLysAlaGluLysIleIleSerPhe	41
Qy	1013	TCTAGAGACTTCGGTATTCAAAACATTTGGATCTTCAGAAAAAGATCAAAATCTATTTAGT	1072
Db	42	-----LeuAspSerGlnAsnIleAsnLysSerAenSerTh	53
Qy	1073	GATAAAGCTTTAAGCTATATACTAAGCAGATACGAGCGCTCTCTCTAAAAATTAACGATCATG	1132
Db	53	rIleLeuLeu-----AsnProLeuAenGluAen-----	62
Qy	1133	AGCTGGCGTTATTGTGGGATTGTTGTCAGAAACAGCGCTAGATTCCTACGGAA-----	1187
Db	63	-----GlyAsnIlePheAsnLeuGluIleLysThrLysLeuAspLeuAlaThrLysIleAl	81
Qy	1188	-ACTCTAAACGAAACG-----CAAAATTTTGGTCA-----AAGGT	1221
Db	81	aSerLeuAenThrAlaIleGlnAsnAsnPheSerAsnIleArgMetThrAenTySerIle	101
Qy	1222	AAGCAGCAAACTATCAGAGAAAATCGGTTATCAGCGCACCATCGGCTTTTAGGAGCTTT	1281
Db	101	eSerAenGluGluAlaGlnLysLeuIlePheAsnAlaIleLeuSerValGlyIleAlaLe	121
Qy	1282	GGCAATCATCTTGCTCTATGTGAGTTGGCTTTGCGCTTGAATGGCAATATGCTTTCAGTGGCGT	1341
Db	121	uIlePheValThrIlePheThrLeuIleArgPheLysTrpThrPheSerLeuAlaIleIle	141
Qy	1342	ATCGCTCTTAAATCATGACCTTTGGCTACCTGTGCGACTGTGTGTTTATGACACATTTCTT	1401
Db	141	ePheSerLeuLeuPheAsnValLeuMetValLeuLeuAlaIleIleThr-----	158
Qy	1402	TTTGAAGAAAATTCAAATAGATTGTCAGGCCATTGGTGCTTTAATGACTGTTATGGGTA	1461
Db	159	-----ArgIleGluIleSerGlnAsnLeuValValAlaIleLeuThrLeuIleGlyTy	176
Qy	1462	TTCATTAACAATACATTGATCATTTTTCATCGTATTGCT-----GAAGATCG	1509
Db	176	rThrValAsnAspThrIleValIlePheAspArgValLysAlaArgPheSerGluIleAs	196
Qy	1510	CCAAGCGAACCTG-----TTTACCCCTATCATGTTTATTAGTATGATGCCCTTCAAAA	1563
Db	196	nHisGluAenValTyLysPheAspLysIleLysGluIleSerLeuGlnAlaIleArgGI	216
Qy	1564	GAGCTTCAGCCGACGGTAATGACACAGCTACAACCTATCATCGTTTGTGTTAATGCTTTT	1623
Db	216	uThrAlaLysArgSerValTyThr-SerLeuThrThrIleLeuThrIleValValLeuMe	236
Qy	1624	GTTTATAGGGCGCTCTGCTCTTTAATTTTGCATTTATTATGACCATAGGATCTTCT	1683
Db	236	tIlePheTyrgLuserIleAspIleValPheSerLeuThrMetLeuIleGlyValIleIle	256
Qy	1684	AGAACTTTTATCGTCTCTTTATATGACCAACCTCTGTGTTGTTTATGGTCCGCTAAAGA	1743
Db	256	eGlyThrTySerSerLeuPheIleAlaThrArgIleTrpIleIleLeuGluSerSerAr	276
Qy	1744	AAATCGC	1750
Db	276	qAsnArg	278

RESULT 28

AAU63575	AAU63575 standard; Protein; 418 AA.
XX	XX
XX	AAU63575;
XX	XX
XX	27-FEB-2002 (first entry)
XX	XX
DE	Propionibacterium acnes immunogenic protein #24471.
XX	XX
XX	SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW	uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW	inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW	dermatological; osteopathic; neuroprotectant.
XX	XX
OS	Propionibacterium acnes.
XX	XX
PN	WO200181581-A2.
XX	XX
PD	01-NOV-2001.
XX	XX
XX	20-APR-2001; 2001WO-US12865.
XX	XX
PR	21-APR-2000; 2000US-199047P.
PR	02-JUN-2000; 2000US-208841P.
PR	07-JUL-2000; 2000US-216747P.
XX	XX
PA	(CORI-) CORIXA CORP.
XX	XX
PI	Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI	L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX	XX
DR	WPI; 2001-616774/71.
DR	N-PSDB; AAS59634.
XX	XX
PT	Propionibacterium acnes polypeptides and nucleic acids useful for
PT	vaccinating against and diagnosing infections, especially useful for
PT	treating acne vulgaris -
XX	XX
PS	Claim 3; SEQ ID NO 24770; 1069pp; English.

PS Claim 3: SEQ ID No 24770: 1069pp: English.

Sequences AU39105-AU68017 represent *Propionibacterium* acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hystertosis and osteomyelitis), uveitis and endophthalmitis. *P. acnes* is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of *P. acnes* in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for *P. acnes* proteins. These antibodies can be used to downregulate expression and activity of *P. acnes* polypeptides and therefore treat *P. acnes* infections. The antibodies may also be used as diagnostic agents for determining *P. acnes* presence, for example, by enzyme linked immunosorbent assay (ELISA).

CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published/pct/sequences](http://ftp.wipo.int/pub/published/pct/sequences).

XX Sequence 418 AA:

Alignment Scores:	4.85e-17	Length:	418
Pred. No.:	234.00	Matches:	77
Score:	234.00	Conservative:	75
Percent Similarity:	46.71%	Mismatches:	121
Best Local Similarity:	23.69%	Indels:	52
Query Match:	7.07%	Gaps:	13
DB:	22		

US-09-868-987-1 (1-1864) x AAU53575 (1-418)

```

QY 788 ATAAAGCATGATTTCTTGAGAGAGATGCAAAAACCTTGGGCTGTTCTGCAAGCTTTT 847
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 66 11eserYrGlyPhe11eGlnArgArglyValTrpYrThr11eThrAlaValAl1le 85
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 848 CTTTAGTTCGCTGCTCGGGTTTGAGACCTGGAATTCGTTTGGGCAATGATTT 907
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 86 11e11eserPhe11eGlyLeuValArgGlyLeuSnn-----LeuGly11eGlnPhe 103
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 908 AAAGAGGGATGCTTACCTTAATCCAAAAGAGCATGCGATCGAGATGGCTCAA 967
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 104 LysGlyValAlGlnPheArg---AlaProAlaHsValAlaThrSerThr11eAspAsp 122
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 968 ATGGGTGGCAAGTTTGCAATAACTACAGAGACCTGCTTTCTTACAGAC--TTC 1024
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 123 ValArgLysSerValLeu-----SerSerGlyAlaProAspMetAspAlaThr 138
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 1025 CGTATTCAACATTTGATCTTCAGAAAAGATCAAAATCTATTAGATTAAGCTTTA 1084
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 139 GluValValSerLeuGlySerAsp-----AlaValGlnValGlnThrArgAlaLeu 155
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 1085 AGC-----TATACATGACGATACGAGCTCTCTCTTAAATTAATGATCATGAGCTG 1138
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 156 AsnAsnAspGlnThrTrpLysValGlnGluAla11eAla11eAlaThr----- 171
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 1139 CGTATTGTGGATGTGTGTCAGAAACAGGCTAGATTCTCTACGAACTCTAAACGA 1198
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 172 -----GlyThrLysMet 175
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 1199 AAGCAAAATTTTGTCTCAAGGTAGACGCAAACTATGCAAGAAATGCTTACAGCG 1258
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 176 SerSerValThrYrSerLysValGlySerGlnTrpGlyGlnThrGlyLysAla 195
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 1259 ACCATGGGCTTTAGAGAGCTTGGCAATCATCTGCTCTAGAGTTGCGCTT-- 1315
    |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 196 11eLysAlaLeuValAlaPheLeuValLeuValMetLeuGln11eTrpAla11ePheArg 215
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 1316 GAATGCAATATGCTTTCAGTCCGCTATGCGCTTAAATCATGACCTTTGGTACCCTG 1375
    |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 216 HistTrpLysMetSer11eAlaLeuValAlaLeuLeuH1sAspLeu11eValThr11e 235
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 1376 GCACTCTGTTATAGCAGCATTTCTTTGGAAGAAATTCAAATAGATTGCCAACCATT 1435
    |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 236 Gly11eTrpAlaLeuValGlyPheThrValSerProSerThrVal-----11e 251
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 1336 GGCTCTTAATGACGTATGCGGATTCATTAACAATATCTTGATCATTTTGATCGT 1495
    |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 252 GlyValLeu---Thr11eLeuGlyTrpSerLeuTrpAspThrValAlaPheAspLys 270
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 1496 ATTGCTGAA-----GATGCCAAGCAAGCACTGTTTACCCCTATGCAATGTT 1540
    |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 271 ValArgGluAsnValAlaGlnAsp11eAspLysArgAspYrThrPheAlaGluGly----- 288
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 1541 TTAATTAATGATGCGCTTCAAAAAGACGTTCAAGCCGACGAGTAAAGACAGCTACACT 1600
    |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 289 ---AlaAsnArgAlaValAlaGlnValLeuValArgSer11eAsnThrThr11eValGly 307
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 1601 CTATCAGTTTGTTAATGCTTTGTTT-----ATAGGCGGCTCCTCGTCTC 1645
    |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 308 ValLeuProValAlaAlaLeuLeuPheAlaGlyAlaPheValLeuGlySerGlyProLeu 327
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 1646 TTTATTATTGATTATTATGACCATAGGATTTCTTTCAGAACTTATGCTCTTTAT 1705
    |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 328 GluAspLeuGlyLeuAlaLeuPheValGlyMet11eValGlyAla11eYrSerSer11ePhe 347
    |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 1706 ATTGCACCACTCTG 1720
    |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 348 11eAlaThrProVal 352
    |||  |||  |||  |||  |||  |||  |||  |||  |||

```

RESULT 29  
AAG91574  
ID AAG91574 standard; Protein; 403 AA.

```

XX AC AAG91574;
XX XX
XX 26-SEP-2001 (first entry)
XX DT
XX C glutamicum protein fragment SEQ ID NO: 5328.
XX DE
XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX OS Corynebacterium glutamicum.
XX PN EP1108790-A2.
XX PD 20-JUN-2001.
XX PF 18-DEC-2000; 2000EP-0127688.
XX PR 16-DEC-1999; 99JP-0377484.
XX PR 07-APR-2000; 2000JP-0159162.
XX PR 03-AUG-2000; 2000JP-0280988.
XX PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senon A, Ikeda M, Ozaki A;
XX DR N-PSDB; AAH6793.
XX DR WPI; 2001-376931/40.
XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX PS Claim 17; SEQ ID NO: 5328; 246pp + Sequence Listing; English.
XX XX
XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a protein described
XX in the exemplification of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office.
XX CC
XX SQ Sequence 403 AA;
XX XX
XX Alignment Scores:
XX Pred. No.: 2,28e-16 Length: 403
XX Score: 228.00 Matches: 72
XX Percent Similarity: 44.97% Conservative: 80
XX Best Local Similarity: 21.30% Mismatches: 132
XX Query Match: 6.88% Indels: 54
XX DB: 22 Gaps: 8
XX
XX US-09-868-987-1 (1-1864) x AAG91574 (1-403)
XX
XX QY 797 GATTTCTTGAGAGATGCAAAAACCTTGGGCTGTTCTGAGATGTTTCTTTAGGT 856
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34 AspPhe11eAlaYrThrLysLeuTrpYrTrp11eThrGly11eLeuLeuVal11eSer 53
XX
XX QY 857 TGGCTTCTCGGCTTTGGAGCCTGGAATTCGTTTGGGAATGATTTTAAAGAGG 916
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 54 11eLeuPhe11eAla11eArgGlyPheSer-----LeuSer11eAspPheGlnGly 71
XX
XX QY 917 TATGCTTACCTTATTCAAAAGAGCATGCGATGAGGATGTGCTCAATGCGTGGC 976
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```
Db 72 ThrLysMetSerMetProAlaSerAspTyrSerThrGluGlnValGluGluThr----- 89
Qy 977 AAAGTTGGCATAAAGTACAGAGAGCTGGTCTTCTTCTAGAGAGCTTCGCGTATTCAACA 1036
Db 90 -----PheThrGluAlaThrGlyLeuThrProGluLeuValGlnLeuValGly 105
Qy 1037 TTTCGATCTTCAGAAAGATCAAAATCTATTTTAGTGTAGTAAGCTTTAAGCTATATAAG 1096
Db 106 SerGlyAspAlaAargThrLeuGluLeuTyr-----SerGluAargLeuSerAspGluAsp 123
Qy 1097 CAGATACAGAGCTCTCTCTCTAAATTAACGATCATGAGCTGGCGTATTGGGATGTT 1156
Db 124 ValGluLysAlaAargLeu-----AlaIleTyr 132
Qy 1157 GTCAGAAACAGGCTAGATTCTCTACGAAACTCTAAACGAAACGCAAAATTTGGTCA 1216
Db 133 GluGluTyrGlnProLeuAasnSerGluGlnProSerProAspAlaIleGlyAasnSer 152
Qy 1217 AAGCTAAGCAGCAAACTACGAAGAAATCGTTATCAGGCGACCATCGGCTTTAGGA 1276
Db 153 ThrValSerGluSerTrpGlySerThrIleThrGlnArgMetValLeuAlaLeuIleAla 172
Qy 1277 GCTTTGGCAATCATCTGCTCTATGTGAGTTGGCTTTGAATGCAATATGCTTTCACT 1336
Db 173 PheLeuValIleAlaAlaIleTyrIleAlaPheAargLeuGluAargGluMetAlaIleAla 192
Qy 1337 CCGGTATGCGCTTTAATTCATGACCTTTTGCTACCTGTGACGTCTGTTTATAGCACAT 1396
Db 193 AlaMetAlaAlaLeuValAsp-----GlyIleValIleAlaGlyIle 207
Qy 1397 TCTTTTGGAGAAATCAATAGATTGCAAGCATGCTGCTTTAATGACTGTTATG 1456
Db 208 TyrAlaValIleGlyLeuGluValSerProAlaThrValIleGlyLeuLeuThrValLeu 227
Qy 1457 GGGTATTCATTAAACAATCTTGATCATTTTGTATCGTATTTCGTGAAGAT----- 1507
Db 228 ThrPheSerIleTyrAspThrValValPheAspLysValAargGluAasnThrGluGly 247
Qy 1508 -----CGCCAAGCGAACTGTTTACCCCTATGCAT 1537
Db 248 PheGluGlySerArgArgThrTyrAlaGluGlnAlaAasnLeu----- 262
Qy 1538 GTTTTAGTAATGATGCCCTTCAAAAGACGTTACGCCACGGAATGACAAACAGCTTACA 1597
Db 263 -----AlaValAasnGlnThrPheMetArgSerIleSerThrIleIle 277
Qy 1598 ACTCTATCAGTTTGTAAATGTTTGTATATA-----GGCGGCTCCTCT 1642
Db 278 SerAlaLeuProIleIleAlaLeuMetValValAlaValTrpMetMetGlyValGlyThr 297
Qy 1643 GTCTTTAATTTTGCATTTATTATGACCATAGGATTCCTTCTAGGAACCTTTATCGTCTCT 1702
Db 298 LeuLysAspLeuAlaLeuIleGlnLeuIleGlyValIleGlyThrPheSerSerVal 317
Qy 1703 TATATTGACACACCTCTGTTGTTTATGTCGCTGCGTAAAGAAATCGCTCAAAA 1756
Db 318 PheLeuAlaThrProLeuLeuValSerLeuLysAsnArgLeuSerLysThrLys 335

RESULT 30
ID AAE15863 standard; Protein; 403 AA.
AC AAE15863;
XX
DT 26-MAR-2002 (first entry)
XX
DE Corynebacterium glutamicum secF protein.
XX
KW Genetically modified bacterial strain; secD; secF; reporter system;
XX enhanced secretion activity.
XX
OS Corynebacterium glutamicum.
XX
```

```
PN WO200185967-A2.
XX
PD 15-NOV-2001.
XX
PF 26-APR-2001; 2001WO-EP04703.
XX
PR 12-MAY-2000; 2000EP-0110021.
XX
PA (DEGS ) DEGUSSA AG.
XX
PI Berens S, Kalinowski J, Puehler A;
XX
DR WPI; 2002-082901/11.
XX
DR N-PSDB; AAD25591.
XX
PT Genetically modified Corynebacterium with enhanced secretion activity
PT useful for production of desired substance e.g. protein, comprises a
PT modification at one of the genes secD and secF
XX
PS Claim 9; Page 41-42; 42pp; English.
XX
CC The present invention relates to genetically modified bacterial strain
CC Corynebacterium glutamicum, comprising a genetical modification at one of
CC the genes secD and secF. The genetically modified bacterial strain is
CC useful for production of desired substance which is an amino acid,
CC oligopeptide, polypeptide or protein preferably a heterologous protein,
CC where the produced substance is secreted by the bacterial strain. The
CC invention is useful in a reporter system. Modification in SecD and SecF
CC in genetically modified bacterial strain Corynebacterium glutamicum,
CC results in enhanced secretion of the strain, which is utilised for
CC production of high amounts of desired substances which can be easily
CC isolated from the source of production. The present sequence is
CC Corynebacterium glutamicum secF protein.
XX
SQ Sequence 403 AA;

Alignment Scores:
Pred. No.: 2,28e-16 Length: 403
Score: 228.00 Matches: 72
Percent Similarity: 44.97% Conservative: 80
Best Local Similarity: 21.30% Mismatches: 132
Query Match: 6.88% Indels: 54
DB: 23 Gaps: 8

US-09-868-987-1 (1-1864) x AAE15863 (1-403)
Qy 797 GATTCTTGAGAGATCGCAAAACCTTTGGCTCTTCTCGAAGTGTCTTTCTTTTAGGT 856
Db 34 AspPheIleAlaLysThrLysLeuTrpTyrTrpIleThrGlyIleLeuLeuValIleSer 53
Qy 857 TGCCTTCTCTCGGGTTTGGAGGCTCGAAATTCGCTTTTGGGAATGGAATTTAAAGGAGGG 916
Db 54 IleLeuPheIleAlaIleAargGlyPheSer-----LeuSerIleAspPheGlnGly 71
Qy 917 TAGCCTTTTACCTTTAATCCAAAGACATGGCATCAGCATGTTGCTCAATCGGTGGC 976
Db 72 ThrLysMetSerMetProAlaSerAspTyrSerThrGluGlnValGluGluThr----- 89
Qy 977 AAAGTTGGCATAAAGTACAGGAAGCTGTTCTTCTCTAGAGACTTCGCTATTCAACA 1036
Db 90 -----PheThrGluAlaThrGlyIleThrProGluLeuValGlnLeuValGly 105
Qy 1037 TTTCGATCTTCAGAAAGATCAAAATCTATTTTAGTGTAGTAAGCTTTAAGCTATATAAG 1096
Db 106 SerGlyAspAlaAargThrLeuGluLeuTyr-----SerGluAargLeuSerAspGluAsp 123
Qy 1097 CAGATACAGAGCTCTCTCTCTAAATTAACGATCATGAGCTGGCGTATTGGGATGTT 1156
Db 124 ValGluLysAlaAargLeu-----AlaIleTyr 132
Qy 1157 GTCAGAAACAGGCTAGATTCTCTACGAAACTCTAAACGAAACGCAAAATTTGGTCA 1216
Db 133 GluGluTyrGlnProLeuAasnSerGluGlnProSerProAspAlaIleGlyAasnSer 152
```

QY 1217 AAGTAAGACGAACTATCAAGAAATGCGTTATCAGGACCATCGGCTTTAGGA 1276  
 DB 153 ThrValSerIuSerTyrGlySerThrIleThrGlnMetValIleValIleLeu 172  
 QY 1277 GCTTGGCAATCATCTTCATGATGAGTTTGCGCTTGAATGGAATGATGCTTTCGT 1336  
 DB 173 PheLeuValIleAlaIleIleTyrIleAlaPheArgLeuGlnArgIleAlaIleAla 192  
 QY 1337 GCCGTATGCGCTTAAATCATGATCCTTTGGCTACCTGTGACCTTGTATAGACAT 1396  
 DB 193 AlameValIleAlaLeuValIleValAsp-----GlyIleValIleAlaGlyIle 207  
 QY 1397 TTCCTTTGAAAGAAATTCATATGATTTGCAAGCATGCTGCTTATATGCTATG 1456  
 DB 208 TyrAlaValIleGlyLeuGlnValSerProAlaThrValIleGlyLeuThrValLeu 227  
 QY 1457 GGGTATTCATTAACATACCTTTGATCATTTTGTATGCTATTCGTAAGAT----- 1507  
 DB 228 ThrPheSerIleTyrAspThrValValAlaPheAspIysValArgIleAsnThrGly 247  
 QY 1508 -----CGCCAAAGCACTGCTTTTACCCCTATGCAT 1537  
 DB 248 PheGlnGlySerArgArgGlyThrTyrAlaGlnGlnAlaAsnLeu----- 262  
 QY 1538 GTTTTATGTAATGATGCCCTTCAAAGCGTTACGCCGAGGTAATGACAACAGCTACA 1597  
 DB 263 -----AlaValAsnGlnThrPheMetArgSerIleSerThrIleIle 277  
 QY 1598 ACTGATCATGTTTGTATGATGCTTTGTTTATA-----GGCGGCTCCTCT 1642  
 DB 278 SerAlaLeuProIleIleAlaLeuMetValAlaValAlaValIlePheMetMetGlyValGlyThr 297  
 QY 1643 GCTTTAATTTTGCATTTATATATGACCATAGGATTTCTTCAAGAACTTATGCTGCTT 1702  
 DB 298 LeuIysAspLeuAlaLeuIleGlnLeuIleGlyValIleGlnGlyThrPheSerSerVal 317  
 QY 1703 TATATGCAACCACTCTGCTGTGTTTATGCTTCAAGAAATCGCTCAAAA 1756  
 DB 318 PheLeuAlaThrProLeuAlaValSerLeuIysAsnArgLeuSerIysThrIys 335

RESULT 31

AAU36596 standard; Protein: 1055 AA.

AAU36596;

14-FEB-2002 (first entry)

Staphylococcus aureus cellular proliferation protein #766.

Antisense; prokaryotic cellular proliferation protein;

antibiotic; antibacterial; drug design.

Staphylococcus aureus.

W0200170955-A2.

27-SEP-2001.

21-MAR-2001; 2001WO-US09180.

21-MAR-2000; 2000US-191078P.

23-MAY-2000; 2000US-206848P.

26-MAY-2000; 2000US-207727P.

23-OCT-2000; 2000US-242578P.

27-NOV-2000; 2000US-253625P.

22-DEC-2000; 2000US-257931P.

16-FEB-2001; 2001US-269308P.

(ELIT-) ELITRA PHARM INC.

Haselebeck R, Ohlsen KU, Zysek JD, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;  
 XX WPI, 2001-611495/70.  
 DR N-PSDB; AAS54455.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Example 3; Seq ID No 12189; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 1055 AA;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	5.18e-07	1055	132	111	229	28
Percent Similarity:	147.50					
Best Local Similarity:	34.42%					
Query Match:	18.70%					
DB:	4.45%					
	22					
	28					

US-09-868-987-1 (1-1864) x AAU36596 (1-1055)

QY 265 GAAAGAAATGATGACCAAGCATTTATCTCAGCATGCTGCTGGCAATGCTTATGT 324  
 DB 366 GluIysSerLeuTyrThrMet-ValGluYsAlaSerLeuGlyThrIleValAlaIleI 385  
 QY 325 TTGATGAGCGTATATTATAGA-----TTTGAGGCGCTATGCTTGGGACCTGCTT 378  
 DB 385 eValIleLeuLeuPheLeuArgAsnIleArgThrAlaIleSerIleIleSerIlePr 405  
 QY 379 TCGAATCTTTTGTATATCTGCGGACGCTTACGATTTTGGATGCGGCACTGCTTGT 438  
 DB 405 OLeuSerLeuLeuMetAlaLeuIleAlaLeuIysLeuSerAspValSerLeuAsnIleI 425  
 QY 439 AGACGCTGCGGATGCTTGTGATGCGGATGCGGCGATGCAATGCTTGTAT 498  
 DB 425 uThrLeuGlyAlaLeuThrValAlaIleGlyArgValIleAspAspSerIleValI 445  
 QY 499 CGAAGAAATC-----CGAAGAAATTTTATGCTTCAAGTCTTAA 540  
 DB 445 IGlAsnIleTyrArgArgLeuThrAspSerGlnGlnLeuIleGlyGluAsnLeu-- 464  
 QY 541 AAAATCTGTGAAAGAAAGATATACCAAGGCTTTTGGAGCATTTTGTATCTTACCTGAC 600  
 DB 465 -----IleIleSerAlaThrThrGlnValPheIysProIleMetSerThrIle 482  
 QY 601 TACAGTATGCGCTCAGCATTTCTTTCTTCAGATATACGAGGCTATTT----- 649  
 DB 482 IThrIleIleValPheLeuProLeuValPheVal---SerGlySerValIleGluMetP 501  
 QY 650 -AAAGGTTTGTGATGATGATTTTATGAGATTTTCTTCAATGTTTACG----- 700  
 DB 501 eArgProPheAlaLeuAlaIleAlaPheSerLeuLeuAlaSerLeuLeuValSerIleTh 521





PS Claim 7; Pages 1092-1095; 1657pp; French.

XX The present invention relates to the genomic sequence of *Pyrococcus*  
 CC abyssal (see AAF86431 and AAF41223-7) and *P. abyssal* proteins. *P. abyssal* is  
 CC a hyperthermophilic archaeon, which is isolated from deep-sea  
 CC hydrothermal vents. The present sequence is one such *P. abyssal* protein.  
 CC The proteins of the present invention have various potential industrial  
 CC uses, since the proteins are stable at very high temperatures, some up to  
 CC 110 degrees centigrade.  
 CC Note: This patent is in the same patent family as WO200065062, which  
 CC contains additional sequences as shown in AAF99132-AAF99143,  
 CC AAF75903-AAF75920 and AAF66436.

CC Sequence 746 AA;

Alignment Scores:

Pred. No.:	4.9e-07	Length:	746
Score:	147.00	Matches:	119
Percent Similarity:	34.93%	Conservative:	115
Best Local Similarity:	17.76%	Mismatches:	258
Query Match:	4.44%	Indels:	178
DB:	22	Gaps:	24

US-09-868-987-1 (1-1864) x AAF66403 (1-746)

```

QY 23 GAGGGATCAGCGGACTGCTAATGACATATTCGCAACCGTGAGTGGCTATGGCT 82
DB 138 GluAlaArgAsnGlyLeuIleSerGlyAspTyr----- 148
QY 83 GTAGTATTCAGCGTTATATGTCAGCAGCGCTTATTTAAACGTCACATG----- 133
DB 149 -----SerSerThrIleIleIleValIleAsnLeuAsnArgGlu 160
QY 134 AAAATGATGCGAGTGTCTCAGGAAATTAC-----CACGTGAGTGAAGCAAACTC 187
DB 161 LysAsnGlnLysAlaLeuValArgValTyrAsnAspIleGluArgGluIleGluArgThr 180
QY 188 -----GCCTAGATTAAATCTGAGCGAGTCTTTGTTCCGAGGTT 232
DB 181 AsnProGlnGlyValGluValValLeuThrGlyAspLeuGlyIleThrTyrLysIle 200
QY 233 CTCAGTGAAGAGAGATCTTCTGATCTTGGAAGAAAAACAATGACCAAGGATATTC 292
DB 201 Leu-----GluMetLeuGlnAsnAspMetAsnArgThrMetAlaIleSerGlyIleIle 218
QY 293 TCAAGATGCTGTGCTGCGCAATGCTTATGTTGATGAGCGTATATTATAGA----- 346
DB 219 -----ValValLeuIleLeuLeuTyrPheTyrLysSerPro 230
QY 347 -----TTTGAGGCGCTGATCGCTTCGGAGCT 373
DB 231 IleArgMetLeuValProLeuValProLeuIlePheGlyValIleMetThrLeuGlyPhe 250
QY 374 GTTCTTCGATCTTTTGTCTTATCTGGCAGCTCTACAGTATTGATGCGCCACTACCC 433
DB 251 MetGlyLeu-----LeuGlyIleProLeuAsp 259
QY 434 TTTCAGAGACTCGTGGAGATTGTTCTTGCTATGGGATGGCGGTAGATGCAAAATGTTCTT 493
DB 260 IleAlaThrThrThrValGlyAlaMetIleIleGlyMetGlyIleAspTyrGlyValHis 279
QY 494 GTATTTCGAAGAATCCGAGAGAAATTTTATGTTCTCAAAAGTCTTAAAAATCTGTAGAA 553
DB 280 ValThrAsnArgTyrTyrGluGluArgGlyLysGlyLysSerLeuGluAlaIleGlu 299
QY 554 AAAGATATACAGGCTTTTGAGGCAATTTTATGTTCTAATGATGACAGTATGGCC 613
DB 300 GluAlaIleAlaGluThrGlyLysAlaLeuLeuGlyAlaAlaIleThrThrIleAlaGly 319
QY 614 TCAGCACTCTTTCTTCTCTAGATACAGGCGCTATTTAAAGGTTTCTTGACATTGATT 673
DB 320 PheLeuAlaLeuSerIleSerIleLeuProSerIleuYsArgLeuSerValSerLeuVal 339

```

```

QY 674 TTAGAAATTTCTTCAATGTTTACGGCTTTTTCATGACTAAATTTTCTTCATGCTG 733
DB 340 MetGlyLeuGlyLeuAlaIleAsnAlaValIleValThrProAlaLeuAlaIleLeu 359
QY 734 TGGATGAATAGACCCACATACAGATTGATGATGATGAATTAAGTCTGGGGATTAAG 793
DB 360 -----GluGluGluPheArgGlyLysValMetLys----- 370
QY 794 CATGATTTCTAGAGAGATGCAAAAACTTGGCGCTTTCTGGAAGT----- 841
DB 371 -----LysGluIleIleAlaIleGlyGlyLysSerLysIle 383
QY 842 -----GTTTTCTTTAGGTTGCGTCTCGGGTTTGAGACCTGGAATTCGGTTTG 895
DB 384 SerPheIlePheSerMetLeuGlySalAlaIleLysArgSerProThrPheAlaLeuLeu 403
QY 896 -----GGAATGATTTTAAAGAGGGTATGCCCTTAATCTTTAAT----- 934
DB 404 IleAlaLeuIleIleSerGlyValSerLeuTyrGlyAlaSerLysIleThrThrGluVal 423
QY 935 -----CCAAAGACATGCGATCGCATGCTTGGCT 964
DB 424 ArgLeuGluLysMetIleProThrAspLeuProGluIleGlnAlaLeuSerAspIleArg 443
QY 965 CAATGCGTGGC----- 976
DB 444 SerGluPheGlyGlyGlnAspGluValThrIleLeuIleLysAlaAspValArgAsp 463
QY 977 -----AAAGTTGTG 985
DB 464 ProThrLeuValArgAspIleLeuArgPheGluArgGluIleLysAlaAspSerTyrIle 483
QY 986 CATTAACATACAGAGACGTGCTTTCTTACAGACTTCGATTAACATTTGGA----- 1042
DB 484 AsnAsnValPheGluThr---GlnSerIleAlaAspValValIleGlnLysTyrGlyTyr 502
QY 1043 -----TCTCAGAAAAGATC-----AAATCTATTATTAAGT 1072
DB 503 IleProGlnAspLysGluLysIleSerGluAlaIleGluGlySerSerLeuValSerSer 522
QY 1073 GATTAAGCTTTAAGCTATACATGATGATGAGATGAGAGCTCTCTCTAAATTAACG----- 1126
DB 523 AspTyrSerMetThrIleIleLys---LeuLysGlyAsnPheMetGlyValThrGlnSer 541
QY 1127 -----ATCATGACTGGCGCTTATGCGGATTTGTTGTCGAAAGAGCGCTAGA 1174
DB 542 GluPheAsnArgIleMetGluTyrPheGluArgAlaIle-----GlnArgAlaAsp 558
QY 1175 TTT-----CTTACGGAACTCTTAACGAACCGCAAAATTTTGG 1213
DB 559 PheProGlnGlyValLysValGluLeuAlaGlyLysSerTyrLeuAsnTyrValLeuAsn 578
QY 1214 TCAAGGTAGACGAAACTATGAAAGAAATGCGTTATCAGCGACATCGGGCTTTTA 1273
DB 579 AspLeuValAsnGluGluLeuGlyLys-----IleSerThrIle 591
QY 1274 GAGCTTTGGCAATCATCTTGTCTATGTGAGTTTGGCTTGAATGCAATATGCTTTC 1333
DB 592 GlyThrLeuIleValAlaMetValIlePheAlaIlePheArgArgProThrValSerIle 611
QY 1334 AGTGCCTATGCGCTTAAATTCATGACCTTTTGGTACCTGTCACATCTTTGTTATAGCA 1393
DB 612 AlaMetIleMetProMetPheLeuGlyAlaLeuThrThr-----IleGly 626
QY 1394 CATTTCTTTTGAAGAAATTCAAATAGATTGCAAGCAATGTCCTTAAATGACTGTA 1453
DB 627 TyrMetGlyLeuAlaGlyIleProPheSerGlnThrLeuAlaGlyValValSerMetIle 646
QY 1454 TTGGGGTATTCATTAACAATATCTTGATGATCTTTGATGTGTAATGCAAGATCCGCA 1513
DB 647 ValGlyLeuGlyValAlaAspTyrGlyMetHisIleThrHisArgPheLeuGluGluAsn 666
QY 1514 GCGAAGCTGTTTACCCCTATGATGTTTAAATGATGATGCGCTTCAAAAGAGCTTACG 1573

```



```

Db 667 GluGlyAenLysThrPro-----:::|||||:::
Qy 1574 GCGCGGTAATGACACAGCTACACTCTATCAGTTTGTAAATGCTTTTCTTTATAGGC 1633
Db 663 GlyIlePheValGlyAlaLeuThrThrAlaGlyGlyPheLeuAlaLeuSerAlaGln 702
Qy 1634 GGCTCTCTGCTTTAATTTGCAATTTATGACCATAGGATCTCTTAGGAACTTTA 1693
Db 703 LeuThrAlaIleHisAspPheGlyArgValLeuAlaValGlyIlePheAlaSerMetPhe 722
Qy 1694 TCGTCTCTTATATGACACACCTCTGTTG 1723
Db 723 AlaAlaTyrLeuValThrProAlaIleLeu 732

RESULT 33
ID AAY35183 standard; Protein; 564 AA.
XX
AC AAY35183;
XX
DT 13-SEP-1999 (first entry)
XX
DE C. pneumoniae protein involved in metabolism of nucleic acids.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
XX
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
PA (G8ST ) GENSET.
XX
PI Griffiths R;
XX
DR WPI; 1999-357842/30.
XX
PS Genome sequence of Chlamydia pneumoniae
XX
Page 1041-1042; Disclosure; 1912pp; English.
XX
AAY34584-Y35879 represent the proteins encoded by all the open reading
frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
C. pneumoniae causes respiratory disease such as pneumonia and
bronchitis and is thought to be a contributing factor in heart
disease, sarcoidosis, sinusitis, purulent otitis media, erythema
nodosum or pharyngitis. The polypeptides encoded by the open reading
frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
immunogenic compositions as vaccines. Vectors containing C. pneumoniae
nucleotide sequences can also be used as immunogenic compositions,
especially where the vector directs the expression of a neutralising
epitope of C. pneumoniae.
XX
SQ Sequence 564 AA;

Alignment Scores:
Pred. No.: 7,1e-07 Length: 564
Score: 145.00 Matches: 25
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.38% Indels: 0
DB: 20 Gaps: 0

```

US-09-868-987-1 (1-1864) x AAY35183 (1-564)

```

Qy 1788 CAATATATAAATCTCCTTTGGGACTTTAGTCCCAAGGCCCTGCTGCTATTAATTTATG 1847
Db 1 GlnTyrIysAsnLeuLeuTrpAspPheSerProIysGlyProCysGlyIleLysPheMet 20
Qy 1848 ACAAAATTCAGATAAT 1862
Db 21 ThrAsnSerAspAsn 25
RESULT 34
AAU34738
ID AAU34738 standard; Protein; 1034 AA.
XX
AC AAU34738;
XX
DT 14-FEB-2002 (first entry)
XX
DE E. coli cellular proliferation protein #319.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-289308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlson KL, Zykkind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS52597.
XX
PT New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 10331; 51pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1034 AA;

```

## Alignment Scores:

Pred. No.:	1,28e-06	Length:	1034
Score:	144.00	Matches:	135
Percent Similarity:	31.73%	Conservative:	89
Best Local Similarity:	19.12%	Mismatches:	226
Query Match:	4.35%	Indels:	256
DB:	22	Gaps:	30

US-09-868-987-1 (1-1864) x AAU34738 (1-1034)

```

QY 308 TTGGCAATGCTTATGTTTATGAGCGTATATATAGATTGGAGGCGTCATCGCTCG 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 LeuValMetCyrLeuPheLeuGlnAsnMetCArgAlaThrLeuIleProThrIleAlaVal 372
QY 368 GGAGCTGTTCTTCTGATCTTTGCTTATCGGCGAGCTCTACAGTATTTGGATGCCCA 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 ProValIleLeuLeuGlyThrPheAlaIleLeuAlaIlePheGlyTyrSerIleAsnThr 392
QY 428 CTCACCTTGTACAGACTCGCTGGGATTTGCTGCTATGAGGATGGCCGTAGATCAAT 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 393 LeuThrMetPheGlyMet-----ValLeuAlaIleGlyLeuLeuValAspAspAla 409
QY 488 GTTCTGTATTCGAAGAATCCGAGAGGAATTTTATGTCTCAAGTCTT-----AAA 541
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 410 IleValIleValIleGluAsnVal---GluArgValMetMetGluAspLysLeuProLys 428
QY 542 AAATCTGTAGAAAAGATATACCAAGGCTTTGGAGCCATTTTGAATTTGAATCTGACT 601
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 429 GluAlaThrGluLysSerMetSerGlnIleGlnIleAlaLeuVal-----GlyIleAla 446
QY 602 ACACTATTTGGCTCAGCACTTCTT-----TTCTTCCTCAAGATACAGGCTTAT 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 447 MetValIleSerAlaValAlaPheIleProMetAlaPhePheGlyGlySerThrGlyAlaIle 466
QY 650 ---AAAGGTTGCTTGTGACATGATTTTGAAGAAATTTCTTCAATGTTTACGCTCTT 706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 467 TyrArgGlnPheSerIleThrIleValSerAlaMetAlaLeuSerValLeuValAlaLeu 486
QY 707 TTCATGACT----- 715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 IleLeuThrProAlaLeuCyAlaThrLeuLeuLysProValSerAlaGluHisIleGlu 506
QY 716 ---AAATTTTCTTCATACCCTGATGATGAATAAGACC---CAACATACACATTCGATATG 769
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 507 AsnLysGlyGlyPhePheGlyTyrPheAsnThrThrPheAspHisSerValAsnHisTyr 526
QY 770 ATGAAAT-----AAGTTCGTGGGATTAAGCATGATTTCTTG----- 805
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 527 ThrAsnSerValIleGlyIleLeuGlySerThrGlyArgTyrLeuLeuIleTyrAlaLeu 546
QY 806 -----AGAGATGCAGAAAACCTTGGGCTGTTCTGGAAGTGTTTTCTT----- 850
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 547 IleValAlaGlyMetValIleLeuPheLeuArgLeuProSerSerPheLeuProGluGlu 566
QY 851 -----TTGATTCGCTGCTCGGCTTTGAGGCTGGAAT----- 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 567 AspGlnGlyValPheLeuThrMetIleGlnLeuProAlaGlyAlaThrGlnGluArgThr 586
QY 886 ----- 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 587 GlnLysValIleAspGlnValThrAspTyrTyrLeuLysAsnGluLysAlaAsnValGlu 606
QY 887 -----TCCGTTTGGGAATGATTTTAAAGA----- 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 607 SerValPheThrValAsnGlyPheSerPheSerGlyGlnAlaGlnAsnAlaGlyMetAla 626
QY 913 ----- 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 627 PheValSerLeuLysProTyrGlnGluArgAsnGlyAspGluAsnSerAlaGluAlaVal 646
QY 914 -----GGGTATGCTTATACCTTAT 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 647 IleHisArgAlaLysMetGluLeuGlyLysIleArgAspGlyPheValIleProPheAsn 666

```

```

QY 934 ----- 934
Db 667 MetProAlaIleValIleGluLeuGlyThrAlaThrGlyPheAspPheGluLeuIleAspGln 686
QY 934 ----- 934
Db 687 AlaGlyLeuGlyHisAspAlaLeuThrGlnAlaArgAsnGlnLeuLeuGlyMetAlaIle 706
QY 935 -----CCAAAAGACATGCATCGACGATGTTGCTCAATG 970
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 707 GlnHisProAlaSerLeuValSerValArgProAsnGlyLeuGluAspThrAlaGlnPhe 726
QY 971 CGTGGCAAGTT---GTGATTAACATAACAGGAAGCTGTCTTCTTCTAGACCTTCGT 1027
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 727 LysLeuGluValAspGlnGluLysGlnAlaIleLeuGlyValSerLeuSerAspIleAsn 746
QY 1028 ATTCAAACATTTGATCTTCA----- 1048
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 747 ---GlnThrIleSerThrAlaLeuGlyGlyThrTyrValAlaAsnAspPheIleAspArgGly 765
QY 1049 -----GAAAAGATCAAAATCTATTATTTAGTATAA 1078
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 ArgValLysLeuTyrValGlnAlaAspAlaLysPheArgMetLeuProGluAspVal 785
QY 1079 GCTTAAAGCTATCTAGACGATACGAGCTCTCTCTTA-----AAATTAAGATC 1129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 786 AspLysLeuTyrValArgSerAlaAsnGlyGluMetValProPheSerAlaPheThrThr 805
QY 1130 ATGAGCTG-----CGTTATTTGGGATTT---GTGTGACA 1162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 806 SerHisThrValTyrGlySerProArgLeuGluArgTyrAsnGlyLeuProSerMetGlu 825
QY 1163 AACAGGCTAGATTTCTTACGAAACTTAAACGAACGA---AAATTTGCTCAAG 1219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 826 IleGlnGlyGluAlaIleProGlyThrSerSerGlyAspAlaMetAlaLeuMetGluAsn 845
QY 1220 GTAAAGCAAACTATTCG-----AAGAAATGCGTTTACAGCGC 1258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 846 LeuAlaSerLysLeuProAlaGlyIleGlyTyrAspTyrThrGlyMetSerTyrGlnGlu 865
QY 1259 ACCATCGGG-----CTTTAGAGACTTTGGCAATCATCTTGCCTATGTC 1303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 866 ArgLeuSerGlyAsnGlnAlaProAlaLeuValAlaIleSerPheValValAlaPheLeu 885
QY 1304 AGTTGGCTTGAATGCAATATGCTTTCAGTCCGATAGCCTTAAATTCATGACTT 1363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 886 CysLeuAlaIleLeuTyrGlnSerTyrSerIleProValSerValMetLeuValAlaPro 905
QY 1364 TTGGCTAACCTGTCAGTCTTGTATTATAGACATTTCTTTTGAAGAAATTCAAATAGAT 1423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 906 LeuGlyIleValIleValIleLeuAlaIleThrLeuPheAsnGlnLys-----AsnAsp 923
QY 1424 TTCGAAGCATTGGTCTTAAATGACTGATTTGGGCTATTCATTAAACAATCTTGTATC 1483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 924 ValTyrPheMetValGlyLeuLeuThrThrIleGlyLeuSerAlaLysAsnAlaIleLeu 943
QY 1484 ATTTTGTATCGTATTCGTGAAGATCGCCAAAGCAGACCTGTTTACCCCTATGATGTTTA 1543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 944 IleValGlu-----Phe 947
QY 1544 GTTAATGATGCCCTTCAAAAG-----ACGTTACGCCGACGGTAATGACACAA 1591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 948 AlaLysAspLeuMetGlnLysGlyLysGlyValValGlnAlaThrLeuMetAlaVal 967
QY 1592 GTTAACTATCATGATTTGTTATG-----CTTTGTTTATA----- 1630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 968 ArgMetArgLeuArgProIleLeuMetThrSerLeuAlaPheIleLeuGlyValLeuPro 987
QY 1631 -----GGCGGCTCCTGCTGCTTAAATTTTGCATTTATATGACCATA 1672
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 988 LeuAlaIleSerAsnGlnAlaGlySerGlyAlaGlnAsn-----AlaValGlyIle 1004

```





Qy 1526 -----ACCCCTATGCGATGTTTAAAGTAAATGATGCC 1555  
 Db 650 IleSerSerIleTrpLysAsnSerLysMetProLeuArgLysAlaIleThrGluGly 669  
 Qy 1556 CTTCAAAAGACGTTCCAGCGCAGGTAATGACACAGCT----- 1594  
 Db 670 ValGlyGlnThrGlyGly-----ValIleThrSerAlaGlyLeuIleLeuAlaGlyThr 687  
 Qy 1595 -----ACAACCTATGCGATGTTTAAAGTAAATGCTTTTATAGGCGGCTCCTCT 1642  
 Db 688 PheGlyValLeuThrThrLeuProIleGlnLeuVal----- 700  
 Qy 1643 GTCCTTAATTTGCAATTTATTATGACCATAGGATCTCTTAGGAATTTATCGTCTCTT 1702  
 Db 701 -----GlnPheGlyLeuIleThrAlaIleGlyValLeuLeuAspThr----- 714  
 Qy 1703 TATATTGACCACTCTCTGTTG 1726  
 Db 715 PheIleValArgProPheLeuVal 722  
 RESULT 37  
 AAB96433  
 ID AAB96433 standard; Protein; 507 AA.  
 XX  
 AC AAB96433;  
 XX  
 DT 29-OCT-2001 (first entry)  
 XX  
 DE Putative P. abyssi preprotein translocase subunit #3.  
 XX  
 KW Hyperthermophilic archaeon; hyperthermophilic protein.  
 XX  
 OS Pyrococcus abyssi.  
 XX  
 PN FR2792651-A1.  
 XX  
 PD 27-OCT-2000.  
 XX  
 PF 21-APR-1999; 99FR-0005034.  
 XX  
 PR 21-APR-1999; 99FR-0005034.  
 XX  
 PA (CNRS) CNRS CENT NAT RECH SCI.  
 XX  
 PI (IFRE-) IFREMER INST FR RECH EXPL MER.  
 XX  
 PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;  
 XX  
 PI Querellou J, Weissenbach J, Saurin W, Heilig R;  
 XX  
 DR WPI; 2001-126236/14.  
 XX  
 XX New nucleotide sequences isolated from Pyrococcus abyssi encode  
 PT proteins useful in industry -  
 XX  
 FS Claim 7; Pages 1128-1129; 1657pp; French.  
 XX  
 CC The present invention relates to the genomic sequence of Pyrococcus  
 CC abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is  
 CC a hyperthermophilic archaeon, which is isolated from deep-sea  
 CC hydrothermal vents. The present sequence is one such P. abyssi protein.  
 CC The proteins of the present invention have various potential industrial  
 CC uses, since the proteins are stable at very high temperatures, some up to  
 CC 110 degrees centigrade.  
 CC Note: This patent is in the same patent family as WO2000065062, which  
 CC contains additional sequences as shown in AAB99132-AAB99143,  
 CC AAH75903-AAH75920 and AAG66436.  
 XX  
 SQ Sequence 507 AA;  
 Alignment Scores:  
 Pred. No.: 2.82e-06 Length: 507  
 Score: 139.50 Matches: 47  
 Percent Similarity: 51.46% Conservative: 41

Best Local Similarity: 27.49% Mismatches: 70  
 Query Match: 4.21% Indels: 13  
 DB: 22 Gaps: 5  
 US-09-868-987-1 (1-1864) x AAB96433 (1-507)  
 Qy 197 TTAATACTGGAGGAGTCTTTTGTCCGAGGTTCTCAGTGAAGACAGATCTCTTCT 256  
 Db 318 LeuArgSerGlySerLeuProValLysLeuSerIleGluArgIleAspTyrIleSerPro 337  
 Qy 257 GATCTTCGGAACCAATGTACACAGGATTTATCTCAGCATCGTGTGGCTTCGCAATG 316  
 Db 338 LysLeuGluAspPheLysArgGlnValIleAlaGlyIleAlaLeuLeuVal 357  
 Qy 317 CTTATTCTTTGATGAGCGTATATTATAGA-----TTTGGAGGCGTCATC 361  
 Db 358 ValGlyAlaIleValTyrLeuHisTyrArgLysLeuLysIleAlaIleProValMetPhe 377  
 Qy 362 GCTTCGGAGCTGTTCTTGAATCTTTTGTCTATCTGGGAGCTCTACAGTAT---TTG 418  
 Db 378 ThrSerPheSerGluValLeuIleLeuGlyIleAlaSerIleIleArgTrpAsnLeu 397  
 Qy 419 GATCGGCACCTCACCTGTCTGAGGACTCGCTGGATTGTTCTTGTATGGGATGCGCGTA 478  
 Db 398 AspLeuPro-----SerIleAlaGlyIleIleAlaIleGlyThrGlyVal 413  
 Qy 479 GATCAAAATGTTCTTGTATTCGAAAGAAATCCGAGAGGAATTTTATTGTCCTCAAAGT--- 535  
 Db 414 AspGlnGlnIleValIleThrAspGluLeuLeuGlyAspValThrAlaGlyLysLysArg 433  
 Qy 536 ---CTTAAAAAATCT---GTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGAT 589  
 Db 434 IleValLysArgSerGlyIleLeuLysArgMetGlyArgAlaPheIleIleLeuAla 453  
 Qy 590 TCTAACTTGACTACAGTATTGGCTCAGCAGCTTCTTTCTTCTTAGATACAGGCGCTATT 649  
 Db 454 SerAlaSerThrThrIleValAlaMetSerPheLeuPheLysPhePheValGlyGlyLeu 473  
 Qy 650 AAAGGTTTGTCTTGACATTTAGGAAATTTAGGAATT 682  
 Db 474 ArgGlyPheAlaPheThrThrIleLeuGlyIle 484  
 RESULT 38  
 AAB30367  
 ID AAB30367 standard; Protein; 1098 AA.  
 XX  
 AC AAB30367;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #30358.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 XX  
 XX WO200175067-A2.  
 PN  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 XX  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Drmanac RT, Liu C, Tang YT;  
 PI  
 XX  
 DR WPI; 2001-639362/73.  
 DR  
 XX N-PSDB; AAS94554.  
 XX



Db 744 AlaMetTyrLeuCysGlyPheSerLeuAsnLeu-----SerLeuMet 758  
 Qy 1439 GCTTTAATGACTGATTTGGGTATTCATTAAACAATCTTTGATCATTTTGTATCGTATT 1498  
 Db 759 AlaLeuThrIleAlaThrGlyPheValValAspAlaIleValValLeuGluAsnIle 778  
 Qy 1499 CGTAGATCCCAAGCAAGCTGTTTACCCTATGCTATGCTGTTTGTATGATGAGCCCTT 1558  
 Db 779 AlaArgHisLeuGluAlaGlyMet---LysProLeuGlnAlaAlaLeuGlnGlyThrArg 797  
 Qy 1559 CAAAGAGCTTTCAGCCGACGCTAATGACACAGCTACACTCTA---TCAGTTTGTGTA 1615  
 Db 798 GluValGlyPhe-----ThrValLeuSerMetSerLeuSerLeuValAlaValPheLeu 815  
 Qy 1616 ATGCTTTTGTATTATAGGCG-----GGCTCCTCTGCTTCTTAATTTTGCATTTATTATG 1666  
 Db 816 ProLeuLeuLeuMetGlyGlyLeuProGlyArgLeuLeuArgGluPheAlaValThrLeu 835  
 Qy 1667 ACCATAGGATCTTCTAGGAACCTTATCGTCTTATATATGACACCTCTGTTGTG 1726  
 Db 836 SerValAlaIleGlyIleSerLeuLeuValSerLeuThrLeuThrProMetCysGly 855  
 Qy 1727 TTTATGGTCCGT 1738  
 Db 856 TrpMetLeuLys 859  
 RESULT 39  
 AAY19957  
 ID AAY19957 standard; Protein; 1007 AA.  
 XX  
 AC AAY19957;  
 XX  
 DT 19-JUL-1999 (first entry)  
 XX  
 DE B. burgdorferi antigenic protein, t814.aa.  
 XX  
 KW Antigenic protein; vaccine; Lyme disease; infection; detection.  
 XX  
 OS Borrelia burgdorferi.  
 XX  
 PN W09859071-A1.  
 XX  
 PD 30-DEC-1998.  
 XX  
 PF 18-JUN-1998; 98WO-US12718.  
 XX  
 PR 03-SEP-1997; 97US-0057483.  
 PR 20-JUN-1997; 97US-0050359.  
 PR 22-JUL-1997; 97US-0053344.  
 PR 22-JUL-1997; 97US-0053377.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (MEDI-) MEDIMUNE INC.  
 XX  
 PI Choi GH, Erwin AL, Hanson MS, Lathigra R;  
 XX  
 DR WPI; 1999-189980/16.  
 DR N-PSDB; AAX61654.  
 XX  
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the diagnosis, prevention and treatment of diseases  
 PT caused by Borrelia, particularly Lyme disease  
 XX  
 PS Claim 12; Page 135; 275pp; English.  
 XX  
 CC This sequence represents a Borrelia burgdorferi (Bb) protein of the  
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
 CC can be used in vaccines for eliciting protective antibodies to members of  
 CC the Borrelia genus, particularly for the use against Lyme disease in  
 CC humans and animals. They can be used for preventing or attenuating an  
 CC infection caused by a member of the Borrelia genus. The products can also  
 CC be used for detection of members of the Borrelia genus.

XX SQ Sequence 1007 AA;  
 Alignment Scores: 8.93e-06 Length: 1007  
 Pred. No.: 136.50 Matches: 128  
 Score: 32.97% Conservative: 117  
 Percent Similarity: 17.23% Mismatches: 243  
 Best Local Similarity: 4.12% Indels: 255  
 Query Match: 20 Gaps: 29  
 DB: 29  
 US-09-868-987-1 (1-1864) x AAY19957 (1-1007)  
 Qy 173 GAAGTGAACAACTC-----GCCTCAGATTAAATCTGGAGCGATCTCTTT 220  
 Db 281 GluIleGluLysLeuLysLeuSerMetProLysAspMetLysLeu----- 295  
 Qy 221 GTTCCCGAGGTTCTCAGTGAAGAGACGATCTCTCTCATCTTGGGAAAAAACAATGTACA 280  
 Db 296 -----GluIleAlaSerAsp-----SerThrAspPheIleLysAlaSerIleSer 310  
 Qy 281 CAAGGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTATTGTTTTCATCAGCGTATAT 340  
 Db 311 ThrValValAsnSerAlaTyrPheGlyAlaMetLeuAlaIlePheValIlePhePhePhe 330  
 Qy 341 TATAGA---TTTGAGGCGTCATCGTTCCGGA---GCTGTCTTCTGAATCTTTTCTGCT 394  
 Db 331 LeuArgSerPheArgAlaThrIleIleGlyIleSerIleProIleAlaIleValLeu 350  
 Qy 395 ATCTGGCAGCTCTACAGTATTGGATGCCCACTACCTTGTCCAGGACCTCGCTGGGATT 454  
 Db 351 ThrPheCysLeuMetTyrPheValAsnIleSerLeuAsnIleMetSerLeuAlaGlyLeu 370  
 Qy 455 GTTCTTCTGATGGGCGCGTAGATGCAAAATGTTCTGTATTTCGAAAGAATCCGAGAG 514  
 Db 371 AlaLeuGlyIleGlyMetValValAspCysSerIleValIleAspAsnIleTyrLys 390  
 Qy 515 GAATTTTATGTTCTCAAGCTTTAAAAATCTGTAGAAAAAGATATACCAAGCGCTTTT 574  
 Db 391 TyrArgGlnLysGlyAlaLysLeuIleSerSerIleLeuGlyAlaGlnGluMetMet 410  
 Qy 575 GGAGCCATTTTGTATTCTTAACCTTGACTACA-----GTATTGGCTCAGCACTTCTTTTC 628  
 Db 411 LeuProIleThrSerSerThrPheThrSerIleCysValPheGlyProPheLeuIlePhe 430  
 Qy 629 TTCCTAGATACAGGCGCT-----ATTAAGCGTTTGTCTTTGACATTGATTTTA 676  
 Db 431 LysSerGluLeuGlyValTyrGlyAspPheLysAspPheThrPheThrIleValIle 450  
 Qy 677 GGAATTTTCTTCAATGTTTTCAGGCTCTTTTCATGACTAAATTTTCTTCATCTGTGG 736  
 Db 451 SerLeuGlyValSerLeuLeuValAlaIlePheLeuValProValLeuSerSerHisTyr 470  
 Qy 737 ATGAATAAGACCCCAACATACACAGTTCATATGATGAATAGTTCTGTGGG----- 787  
 Db 471 ValGlyLeuTyrThrSerPheGlnLysAsnIleLysAsnAlaPheIleArgLysIleAsp 490  
 Qy 787 ----- 787  
 Db 491 AlaPhePheAlaSerIleTyrTyrPheLeuGluPheLeuTyrIleAsnLeuLeuAsnIle 510  
 Qy 788 ---ATAAGCATGATTTCTTGGAGAGGATGCAAAAAACTTTTGGGCTGTTTCTGGAAGTGT 844  
 Db 511 ValLeuAsnHisLysLeuIlePheGlyLeuIleValPhePheSerPheIleGlySerLeu 530  
 Qy 845 -----TTTCTTTAGGTTGGCTCTCGGGTTTGGAGCCTGGAAATTCGTT--- 892  
 Db 531 LeuLeuGlyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 550  
 Qy 893 TTGGGAATGATTTT----- 907  
 Db 551 IleAsnLeuAsnPheProHisLysThrAsnLeuGluTyrAlaLysPheTyrSerAsnArg 570

```

QY 908 -----AAAGAGCGTAT----- 919
Db 571 PheLeuGluIleValIysSerGluAlaIysGlyTyrIysSerIleIleAlaThrLeuArg 590
QY 920 -----GCCTTACCTTAAATCCA-----AAAGACATGGCATCAGCATGTT 961
Db 591 AlaAspArgIleThrPheAsnValLeuPheProLeuIysGluIleuSerArgAspAsnLeu 610
QY 962 GCTCA-----ATGCGTGCAAGTGTGCATTAACCTACAGGAAGCT 1003
Db 611 ThrGlnSerValAspTyrAspSerIleIleTyrIleMetAsnArgIle----- 627
QY 1004 GGTCTTCTTACAGACTTCCGATTCGA-----ACATTGCACT 1045
Db 628 GlyAsnLeuTyrProGluPheAsnIleGluProSerIleSerGlyAsnAlaLeuGlyGly 647
QY 1046 TCAGAAAAGATCAAAATATATTAGTATAGCTTAAGCTTACTACTAAG----- 1096
Db 648 GlyAspSerIleIleValIleValIleSerAlaAsnAspPheGluTyrIleIleAspTyrGly 667
QY 1097 CAGATACGAGCCTCTCTCTAAATTTAAGCATC----- 1129
Db 668 LysIleLeuValSerMetLeuIleValGlyIleProGluLeuValAsnProArgLeuSer 687
QY 1130 -----ATGAGCTGGCGCTTAT 1144
Db 688 IleSerAspPheGlnLeuGlnIleGlyValGluIleAspArgAlaLeuValTyrAsnTyr 707
QY 1145 -----TGCGGATTTGTTGTC 1159
Db 708 GlyIleAspMetAsnThrIleLeuAsnGluLeuIysAlaAsnIleAsnGlyValValAla 727
QY 1159 ----- 1159
Db 728 GlyGlnTyrValGluIysGlyLeuAsnTyrAspIleValLeuIysLeuAspArgMetAsp 747
QY 1159 ----- 1159
Db 748 ValIysAsnLeuIysAspLeuGluIlePheIleThrAsnSerSerGlyValIysIle 767
QY 1160 -----AGA 1162
Db 768 ProPheSerSerIleAlaThrPheGluIysThrAsnIysAlaGluSerIleTyrArgGlu 787
QY 1163 AACAGCGCTAGATTCTCTCGAAGAAC-----CTTAA 1195
Db 788 AsnGlnAlaLeuThrIleTyrLeuAsnIleGlyIleSerProAspAspAsnLeuThrGln 807
QY 1196 CGAAACGCAAAATTTGGTCAAGTAAAGCAGCAAACTATGCAAGAA----- 1243
Db 808 ValThrAlaIysValIasAspPheIleAsnAsnIysValProHisIlyGluGlyIleThr 827
QY 1244 -----ATGCGTTATCAGGCGACCATCGGG 1267
Db 828 LeuIysValGluGlyGluTyrAsnGluPheSerAsnIleMetAsnGlnPheIleIle 847
QY 1268 CTTTAGAGCCTTGGCAATCATCTGCTTATGAGTGTGCGCTTGAATGGCAATAT 1327
Db 848 IleMetAlaIleIleValIleValPheGlyIleMetAlaSerGlnPheGlu----- 864
QY 1328 GCTTTCAGTCCGATGCGCTTAAATCATATCTTTGGCTACTGTCGAGCTTGT 1387
Db 865 SerPheLeuIysProPheIleIleIlePheThrIleProLeuThrAlaIleGlyValVal 884
QY 1388 ATACAGCATTTCTTTTGG-----AGGAATTCGAATGATTTGCAAGCCATGGCTTTA 1444
Db 885 LeuIleHisPheLeuAlaGlyGluIysLeuSerIle-----PheAlaAlaIleIleMetLeu 903
QY 1445 ATGACTGATTTGGGATTCATTAACATATCTTGATCATTTTGTATCGTATCGTGAA 1504
Db 904 MetLeuVal-----GlyValValValAsnThrGlyIleValLeuValAsp----- 918
QY 1505 GATGCCCAAGCAACCTGTTTACCCTTATGCATGTT----- 1540

```

```

Db 919 -----TyrThrGlyLeuLeuIleIleYasArgGlyPheGlyLeuArgGlu 932
QY 1541 TTAGTAAATGATGCTCCCTTCAAAAGACGTTACGCCGATTAATGACACGCTACACT 1600
Db 933 AlaIleIleGluSerCysArgSerArgLeuArgProIleuMetSerSerLeuThrSer 952
QY 1601 CTATCAGTTTGTATAGCTTTTGTATTATAGCGGCTCTCT-----GCTTTAAT 1651
Db 953 IleIleGlyLeuIleProMetAlaPheSerSerGlySerGlyAsnGluLeuIysPro 972
QY 1652 TTTGCAATTTATTAATGACATAGGATTTCTCTAGAACTTATGCTCTTTATATGCA 1711
Db 973 IleAlaPheThrPheIleGlyGlyMetThrAlaSerThrPheLeuThrLeuPheIle 992
QY 1712 CCACCTGTG 1720
Db 993 ProMetLeu 995

RESULT 40
AAU35552
ID AAU35552 standard; Protein; 1032 AA.
AC AAU35552;
XX
DT 14-FEB-2002 (first entry)
XX
DE Haemophilus influenzae cellular proliferation protein #193.
XX
DE Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
OS Haemophilus influenzae.
XX
PN MO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001MO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GT,
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
XX
DR N-FSDB; AAS53411.
XX
PS Example 3; Seq ID No 1145; 511pp; English.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen

```



for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.  
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published](http://wipo.int/pub/published) pct sequences.

Sequence 1032 AA;

Alignment Scores:	
Pred. No.:	9,058-06
Score:	136.50
Percent Similarity:	34.91%
Best Local Similarity:	16.95%
Query Match:	4.12%
DB:	22
	Gaps: 30
	Indels: 283
	Mismatches: 241
	Conservative: 145
	Matches: 136
	Length: 1032

US-09-868-987-1 (1-1864) x AAU35552 (1-1032)

Qy	44	AATGACAATATTCTGCAAAACCGTGGATGCGGTAGCTGTAGTGAATGACGGTTATATG	103
Db	274	AsnSerArgalathAlaAsnGlyAlaGluSerValValLeuAlaIleAsnProThrSer	293
Qy	104	GTACAGCACCCTATTTAAACGTCCTCCCATTCGAAATCATCCAGTGTCTCAGGGAATTT	163
Db	294	ThrAlaAsnProLeu-----ThrValAlaGluLys---	303
Qy	164	ACCCACCGTGAAGTCAGCAAAACTCGCCTCAGATTTAAATCTCGGAGCGATG---	220
Db	304	-----IleArgProLeuIleGluSerIleIleGluSerIleGluSerMet	319
Qy	221	GTTCGCCAGGTCTTCAGTGAAGACAGCATCTCTTCGATCTTGGGAAAAACAATGTACA	280
Db	320	GluSerAspIleLeuIleAspArgThrIleAlaIleAsnSerSerIleHisGluValIle	339
Qy	281	CAAGCCATTATCTCAGCATGCTGTGGCTGGCAATGCTTATTCTTTGGATGACGCTATAT	340
Db	340	LysThrIle-----GlyGluAlaThrLeuIleValLeuValValIleLeu	354
Qy	341	TATAGATTTGGAGC-----GTCAATCGCTTCGGGAGCTGTTCTTCTGAATCTTT	388
Db	355	MetPheIleGlySerPheArgAlaIleLeuIleProIleLeuAlaIleProIleSerLeu	374
Qy	389	TTCCTTATCTGGCAGCTCTACAGTATTTGGATGCGCCACTCACCTGTGCAGACATCGCT	448
Db	375	IleGlyValLeuMetLeuLeuGlnSerPheAsnPheSerIleAsnLeuMetThrLeuLeu	394
Qy	449	GGGATGCTCTTGTATGGGATGGCCGTAGATGCAAACTCTCTGTATTTCGAAAGATC	508
Db	395	AlaLeuIleAlaIleGlyLeuValAspAlaIleValLeuGluGluAsnIle	414
Qy	509	CGAGGAGAAATTTTATTGTCTCAAGTCTTAAAAAATCTGTAGAAAAAGGATATACCAAG	568
Db	415	-----AspArgHisIleLysAlaGlyGluThrPro	424
Qy	569	GCTTTTGGAGCAATTTTGATCTTAAC-----	595
Db	425	PheArgAlaAlaIleIleGlyThrArgGluIleAlaValProValIleSerMetThrIle	444
Qy	596	---TTGACTACGATTTGGCGCTCAGCACTCTTTTCTTCTAGATACAGGCGCTATT---	649
Db	445	AlaLeuIleAlaValIleIleSerProMetAlaLeuMetGlyGlyIleThrGlyThrLeuPhe	464
Qy	650	AAAGGGTTTCCTTGACATGATTTTAGGAATTTCTCTCAATGTTTACGGCTCTTTTC	709
Db	465	LysGluPheAlaLeuThrLeuAlaGlyAlaValPheIleSerGlyValValAlaLeuThr	484
Qy	710	ATGATCAAAATTTTCTTTCATG-----CTGTGGATG---	739
Db	485	LeuSerProMetMetSerSerLysLeuLeuLysSerAsnAlaLysProThrTrpMetGlu	504
Qy	740	AATAAGACCAACATACA-----CAGTTG	763

Db	505	GluArgValGluHisThrLeuGlyLysValAsnArgValTyrgluTyrrMetLeuAspLeu	524
Qy	764	CATATGATGAATAAGTTCGTGGGATAAAGCATGATTCTTTGAGAGGATGCAAAAACCT	823
Db	525	ValMetLeuAsnArg-----LysSerMet	532
Qy	824	TGGGCTGTTTCGGAAGTGTTTCTTTTA-----	853
Db	533	LeuAlaPheAlaValIlePheSerThrLeuProPheLeuPheAsnSerLeuSerSer	552
Qy	854	-----GGTTCGCTTGCTCTCGGGTTTGAGACCTGGAAT	886
Db	553	gluLeuThrProAsnGluAspLysGlyAlaPheIleAlaIleGly---AsnAlaProSer	571
Qy	887	TCCGTT-----	892
Db	572	SerValAsnValAspTyrrIleGlnAsnAlaMetGlnProTyrrMetLysAsnValMetGlu	591
Qy	893	-----TTGGGAATGGATTTTAAAGGAGGATGCTCTTTACCTTTTAATCCA	937
Db	592	ThrProGluValSerPheGlyMetSerIleAlaGly-----AlaProThrSerAsnSer	609
Qy	938	AAAGACATGGCATCAGC-----GATGTTGCTCAAAATCCGTGGCAAAAGTT	982
Db	610	SerLeuAsnIleIleThrLeuLysAspTrpLysGluArgSerArgLysGlnSerAlaIle	629
Qy	983	GTGCATAAATACACGGAAGCTGCTTCTTCTCTAGAGACTTCCGTATTCAACATTT	1039
Db	630	MetAsnGluIleAsnGluLysAlaLysSerIleProGluValSerValSerAlaPheAsn	649
Qy	1039	-----	1039
Db	650	IleProGluIleAspThrGlyGluGlnGlyProProValSerIleValLeuLysThrAla	669
Qy	1040	-----GGATCTTCAGAAAAGATC-----	1057
Db	670	GlnAspTyrrLysSerLeuAlaAsnThrAlaGluLysPheLeuSerAlaMetLysAlaSer	689
Qy	1058	---AAATCTATTATTAGTCATAAAGCTTTAAAGCTAT---ACTAAGCAGATACGAGCCTCT	1111
Db	690	GlyLysPheIleTyrrThrAsnLeuAspLeuThrTyrrAspThrAlaGlnMetThrIleSer	709
Qy	1112	CTCTAAAA-----TTAAGCATCATG-----	1132
Db	710	ValAspLysGluLysAlaGlyThrTyrrGlyIleThrMetGlnGlnIleSerAsnThrLeu	729
Qy	1133	-----AGCTGGCGT	1141
Db	730	GlySerPheLeuSerGlyAlaThrValThrArgValAspValAspGlyArgAlaTyrrLys	749
Qy	1142	TATTGTGGATGTTGCTCAGAAC-----AGCCTAGATTT	1177
Db	750	ValIleSerGlnValLysArgAspAspArgLeuSerProGluSerPheGlnAsnTyrr	769
Qy	1178	CTCTACGAAACTCTAAACGAAACGCAAAATTTGGTCAAAAGGTAAGCAGCAAACTATCG	1237
Db	770	LeuThrAlaSerAsnGlyGlnSerValProLeuSerSerValIleSerMetLysLeuGlu	789
Qy	1238	AAGAAA-----	1243
Db	790	ThrGlnProThrSerLeuProArgPheSerGlnLeuAsnSerAlaGluIleSerAlaVal	809
Qy	1244	-----ATCGGTTATCAGCCGACC---	1261
Db	810	ProMetProGlyIleSerSerGlyAspAlaIleAlaTrpLeuGlnGlnAlaThrAsp	829
Qy	1261	-----	1261
Db	830	AsnLeuProGlnGlyTyrrThrPheAspPheLysSerGluAlaArgGlnLeuValGlnGlu	849
Qy	1262	-----ATCGGGCTTTAGGAGCTTTGGCAATCATCTTGTCTCTATGTGAGTTTCGCG	1312

Db 850 GlySnaIaLeuAlaValThrPheAlaLeuAlaValIleIleIlePheLeuValLeuAla 869  
 QY 1313 TTGAATGGCAATATGCTTTCAGTCCGTATGCGCTTAATCATGACCTTTGGCTACC 1372  
 Db 870 ILeGlnPheGlnSerIleArgAspProMetValIleMetIleSerValProLeuAlaVal 889  
 QY 1373 TGTGCAGCTTG-----TTTATGACACTTTCTTTTGACAAATAATCAATA 1420  
 Db 890 SerGlyAlaLeuValSerLeuAsnIleLeuSerPhePheSerIleAlaGlyThrLeu 909  
 QY 1421 GATTGCAAGCCATTGGTGGCTTTAATGACTGATGGGGTATTCAATTAACAAATCTTGG 1480  
 Db 910 AsnIleTySerGlnValGlyLeuIleThrLeuValGlyLeuIleThrLeuHisGlyIle 929  
 QY 1481 ATCATTTTGTATGCTATTCGTGAAGATCGCAAGCAACTGTTTACCCCTATGCACTTT 1540  
 Db 930 LeuMetCysGluValAlaIalysGluGlnLeuAsnHisGlyLysThrArgIleGluAla 949  
 QY 1541 TTAGTATATGATGCCCTTCACAAAGACGTTCACGCCGAGGTATGACAAACGCTACACT 1600  
 Db 950 IleThrHis--AlaAlaIalysValArgLeuArgProIleLeuMetThrAlaAlaMet 968  
 QY 1601 CTATCAGTTTGTATGCTATTCGTGAAGATCGCAAGCAACTGTTTACCCCTATGCACTTT 1651  
 Db 969 ValAlaGlyLeuIleProLeuLeuTyraIalThrGlyAlaValSerArgPheSer 988  
 QY 1652 TTTGCATTTATATGACCAATAGGATCTTCTAGAACTTATGCTCTTTATATGCA 1711  
 Db 989 IleGlyIleValIleValAlaIalysLeuSerIleGlyThrIlePheThrLeuPheValLeu 1008  
 QY 1712 CCA-----CCTGCTGTTGTTTATGCTCCGT 1738  
 Db 1009 ProValValTySerTyraValAlaThrGlnHisLysProLeuProValPhe----- 1025  
 QY 1739 AAGAAATCGCTCA 1753  
 Db 1026 AspGluAsnLysThr 1030  
 RESULT 41  
 ID AAY19956 standard; Protein, 1036 AA.  
 AC AAY19956;  
 XX  
 DT 19-JUL-1999 (first entry)  
 XX  
 DE B. burgdorferi antigenic protein, f814.aa.  
 XX  
 KM Antigenic protein; vaccine; Lyme disease; infection; detection.  
 XX  
 OS Borrelia burgdorferi.  
 XX  
 PN W09859071-Al.  
 PD 30-DEC-1998.  
 XX  
 PF 18-JUN-1998; 98WO-US12718.  
 PR 03-SEP-1997; 97US-0057483.  
 PR 20-JUN-1997; 97US-0050359.  
 PR 22-JUL-1997; 97US-0053344.  
 PR 22-JUL-1997; 97US-0053377.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (MEDI-) MEDIMUNE INC.  
 PI Choi GH, Erwin AL, Hanson MS, Lachigra R;  
 DR WPI; 1999-189980/16.  
 DR N-PSDB; AAX61653.  
 XX  
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop products for the diagnosis, prevention and treatment of diseases

PT caused by Borrelia, particularly Lyme disease  
 XX  
 PS Claim 12; Page 135; 275bp; English.  
 XX  
 CC This sequence represents a Borrelia burgdorferi (Bb) protein of the  
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
 CC can be used in vaccines for eliciting protective antibodies to members of  
 CC the Borrelia genus, particularly for the use against Lyme disease in  
 CC humans and animals. They can be used for preventing or attenuating an  
 CC infection caused by a member of the Borrelia genus. The products can also  
 CC be used for detection of members of the Borrelia genus.  
 XX  
 SQ Sequence 1036 AA:  
 Alignment Scores:  
 Pred. No.: 9.07e-06 Length: 1036  
 Score: 136.50 Matches: 128  
 Percent Similarity: 32.97% Conserved: 117  
 Best Local Similarity: 17.23% Mismatches: 243  
 Query Match: 4.12% Indels: 255  
 DB: Gaps: 29  
 US-09-868-987-1 (1-1864) x AAY19956 (1-1036)  
 QY 173 GAAGTGAAGCAATC-----GCCCTGATTTAAATCTGGAGGATGCTTTT 220  
 Db 310 GlnIleGlnLysLeuLysLeuSerMetProLysAspMetLysLeu----- 324  
 QY 221 GTTCCAGGTTTCAGTGAAGAGACGATCTCTTGATCTTGGAAAAAATGATACA 280  
 Db 325 -----GlnIleAlaSerAsp-----SerThrAspPheIleLysAlaSerIleSer 339  
 QY 281 CAAGCATTTATCTGACATCTGCTTGGCAATGCTTATGTTTGAAGAGCTATAT 340  
 Db 340 ThrValAlaAsnSerAlaTyraPheGlyAlaMetLeuAlaIlePhePhePhe 359  
 QY 341 TATGCA---TTGGAGGCGTCATCGCTTGGGA---GCTGTTCTTGAAATCTTTGCT 394  
 Db 360 LeuArgSerPheArgAlaThrIleIleIleIleGlyIleSerIleProIleAlaIleValLeu 379  
 QY 395 ATCTGGACAGCTTACAGTATTTGGATGCGCCACTGCTGTCAGAGCTCGTGAGTT 454  
 Db 380 ThrPheCysLeuMetTyraPheValAsnIleSerLeuAsnIleMetSerLeuAlaGlyLeu 399  
 QY 455 GTTCTGCTATGGGAGTGGCGGTAGATCAAAATGTTTGTATTCGAAAGATCCGAG 514  
 Db 400 AlaLeuGlyIleGlyMetValAlaAspCysSerIleValAlaIleAspAsnIleTyraLys 419  
 QY 515 GAATTTTATTTGCTCAAACTCTTAAATAAATCGTGAATAAAGATATACCAAGGCTTTT 574  
 Db 420 TyraGlnLysGlyAlaLysLeuIleSerSerSerIleLeuGlyAlaGlnLysMet 439  
 QY 575 GAGCCATTTTGTATCTTAATCTGACTACA-----GTATGGCTCAGACATCTTTTC 628  
 Db 440 LeuProIleThrSerSerThrPheThrSerIleCysValPheGlyProPheLeuIlePhe 459  
 QY 629 TTCCTAGTACAGGCGCT-----ATTAAGGTTTGCTTGACATGATTTTA 676  
 Db 460 LysSerGlnLeuGlyAlaTyraGlyAspPhePheLysAspPheThrPheThrIleValIle 479  
 QY 677 GGAATTTTCTCTTCATGATTTACGCTCTTTCATGACTAAATTTTCTTCATGCTGTGG 736  
 Db 480 SerLeuGlyValSerLeuLeuValAlaIlePheLeuValProValLeuSerHisTyra 499  
 QY 737 ATGAATAAGACCCCAATACATACAGTTGCATATGATGAATAGTTCGTGGG----- 787  
 Db 500 ValGlyLeuTyraThrSerPheGlnLysAsnIleLysAsnIleAlaGlyIleAsp 519  
 QY 787 ----- 787  
 Db 520 AlaPhePheAlaSerIleTyraTyraPheLeuGlnPheLeuTyraIleAsnLeuLeuAsnIle 539  
 QY 788 ---ATAAAGATGATTTTCTTGAGAGGATGCAAAAAAATTTGGGCTGTTTCTGGAAGCTTT 844

Db 894 SerPheLeuYsProPheIleIleIlePheThrIleProLeuThrAlaIleGlyValVal 913  
 Qy 1388 ATAGCACATTTCTTTTTTG---AAGAAAAATTCAAATAGATTTTGCAGCGCATTTGGTGCTTTA 1444  
 ::: ||||| :::|||::: ||||| ||||| ||||| |||||  
 Db 914 LeuIleHisPheLeuAlaGlyGlulysLeuSerIle---PheAlaAlaIleGlyMetLeu 932  
 Qy 1445 ATGACTGTATTGGGTATTCATTAACAATACTTTTGATCATTTTGTATCGTATTCGTGCGAA 1504  
 ||| ||| ||| :::||| :::||| :::||| :::|||  
 Db 933 MetLeuVal---GlyValValAsnThrGlyIleValLeuValAsp----- 947  
 Qy 1505 GATGCCAACGAACCTGTTTACCCTATGCATGTT----- 1540  
 :::||| :::||| :::||| :::||| :::||| :::||| :::||| :::||| :::||| :::|||  
 Db 948 -----TythrGlyLeuLeuIleLysArgGlyPheGlyLeuArgGlu 961  
 Qy 1541 TTAGTTAATGATGCCCTTCAAAAGACGTTTCAGCCGCACGGTAATAGCAACAGCTACAAC 1600  
 ::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
 Db 962 AlaIleIleGluSerCysArgSerArgLeuArgProIleLeuMetSerSerLeuThrSer 981  
 Qy 1601 CTATCAGTTTTGTTAATGCTTTTGTATAGCGGCTCTCT-----GTCCTTAAT 1651  
 ::: ||||: .::: ||| ||||| :::: :::: :::: :::: :::: ::::  
 Db 982 IleIleGlyLeuIleProMetAlaPheSerSerGlySerGlyAsnGluLeuLeuLysPro 1001  
 Qy 1652 TTTGTCATTATTATGACCATAGGGATCTTCTAGGAACCTTTATCGTCTCTTATATTGCA 1711  
 ||||| ||||| :::: ||| ||||| :::: :::: :::: :::: :::: ::::  
 Db 1002 IleAlaPheThrPheIleGlyMetThrAlaSerThrPheLeuThrLeuPhePheIle 1021  
 Qy 1712 CCACCTCTG 1720  
 ||| |||  
 Db 1022 ProMetLeu 1024  
 ||| |||  
 RESULT 42  
 ABP40160  
 ID ABP40160 standard; Protein; 1051 AA.  
 XX AC ABP40160;  
 XX DT 24-JUL-2002 (first entry)  
 XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5005.  
 XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection  
 XX KW antibacterial; gene therapy.  
 XX OS Staphylococcus epidermidis.  
 XX PN US6380370-B1.  
 XX PD 30-APR-2002.  
 XX PF 13-AUG-1998; 98US-0134001.  
 XX PR 14-AUG-1997; 97US-055779P.  
 XX PR 08-NOV-1997; 97US-064964P.  
 XX PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX PI Doucette-Stamm LA, Bush D;  
 XX WPI; 2002-381255/41.  
 XX DR N-PSDB; ABN92705.  
 Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
 polypeptide, useful for diagnosing and treating bacterial infections -  
 Disclosure; SEQ ID 5005; 267pp; English.  
 ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 frame (ORF) nucleic acid sequences which encode the amino acid sequenc  
 given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 antibacterial activity and can be used in gene therapy. The sequences  
 can also be used in the diagnosis and treatment of bacterial infection  
 particularly S. epidermidis infections. The sequences can be used to  
 screen for compounds able to interfere with the S. epidermidis life



Db 893 ILeValThrLeuLeuValThrPheArgGlyLeuAlaProPheThrIleLeu 912  
 Qy 1343 TCGCTTTAAATTCAGTACCTTTGGCTACCTGTCAGCTGTTTATAGCACATTTCTTT 1402  
 Db 913 PheSerLeuProPheThrValIleGlyValValLeuAlaLeuIleThrGly----- 930  
 Qy 1403 TTGAAGAAAATTCAAATAGATTGCAAGCCATTGTGTCTTTAAATGACTGTATGGGGTAT 1462  
 Db 931 ---GluThrIleSerValPro---SerLeuIleGlyMetLeuMet---LeuIleGlyIle 947  
 Qy 1463 TCATTAAACAATACCTTCATCATTTTTCATCGTATTTCGTAAGATCGCAAGCGAACCTG 1522  
 Db 948 ValValThrAsnAlaIleValLeuIleAspArgValIleAsnAsnGluLeuGlnGlyMet 967  
 Qy 1523 TTTACCCCTATGTCATGTTTATGTTAATGATGCCCTTCCTCAAAAGACGTTTCAGCGCCACGGTA 1582  
 Db 968 -----ProMetLysGlnAlaLeuIleGluAlaGlyThrArgIleArgProIleLeu 985  
 Qy 1583 ATGACACAGCTACACTCTATCACTTTTGTAAATGCTTTTGTATAGCGGGCTCTCT 1642  
 Db 986 MetThrAlaIleAlaThrIleGlyAlaLeuValProLeuLeuPheGlyHisAspSerSer 1005  
 Qy 1643 GTCITT-----AATTGTCATTATTATGACCATAGGATTCCTTAGGACTTTA 1693  
 Db 1006 IleLeuIleSerLysGlyMetAlaAlaThrValIleGlyLeuIleSerSerThrLeu 1025  
 Qy 1694 TCGTCTCTTTATATTGACCACTCTG-----TTGTTGTTTATGTCGCGTAAAGAAAT 1747  
 Db 1026 LeuThrLeuPheValProValIleThrGluIleLeuPheThrLeuIleAsnLysIle 1045  
 Qy 1748 CGCTCAAAA 1756  
 Db 1046 ThrLysLys 1048  
 RESULT 43  
 AAG90442  
 ID AAG90442 standard; Protein; 730 AA.  
 XX  
 AC AAG90442;  
 XX  
 XX 26-SEP-2001 (first entry)  
 XX  
 DE C glutamicum protein fragment SEQ ID NO: 4196.  
 XX  
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 EP1108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-0127688.  
 XX  
 PR 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX  
 DR WPI; 2001-376931/40.  
 DR N-PSDB; AAH65661.  
 XX  
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene

PS Claim 17; SEQ ID NO: 4196; 246pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 SQ Sequence 730 AA;

## Alignment Scores:

Pred. No.: 3.6e-05 Length: 730  
 Score: 130.50 Matches: 128  
 Percent Similarity: 33.55% Conservative: 78  
 Best Local Similarity: 20.85% Mismatches: 270  
 Query Match: 3.94% Indels: 138  
 DB: 22 Gaps: 23

US-09-868-987-1 (1-1864) x AAG90442 (1-730)

Qy 125 GTCCATTGAAAATCATGCCAGTGTCTCAGGGAATTTACCCACCGTGAAGTCAGCAAA 184  
 Db 132 ValProLeuAsnProSerAlaGluLeuThr-----GluSerValGluLys 146  
 Qy 185 CTCGCCTCAGATTTAAATCTGGAGCGATGCTTTTGTTCCTCCGAG---GTTCTCAGTGAA 241  
 Db 147 LeuSerGluThrLeuThrGlnThrProAspThrValSerThrThrValThrGlyPro 166  
 Qy 242 GAGACGATCTCTCTGATCTTTGGGAAAAACAATGT---ACACAAGGCAATTTATCTCACA 298  
 Db 167 AlaGlyPheThrAlaAspLeuSerAlaAlaPheAlaGlyIleAspGlyLeuLeuAla 186  
 Qy 299 TGCTGTGGCTGGCAATGCTTATTGTTTGTAGAGCGTATATTATAGATT--- 349  
 Db 187 ValAlaLeuAlaAlaValLeuValIleLeuValIleValThrArgSerPheIleLeuPro 206  
 Qy 350 GGAGCGTCATCGCTTCGGGA-----GCTGTTCTTGAATCTTTGCTTATCTGGGCA 403  
 Db 207 IleAlaValLeuAlaThrSerLeuPheAlaLeuThrValAlaLeuLeuValThrTrp 226  
 Qy 404 GCTCTACAGTATTGGATGCGCCACTCACCTTGTGACGACTCGCTGGGATT-----GTT 457  
 Db 227 LeuAlaLysTrp---AspIleLeuLeuLeuSerGlyGlnThrGlnGlyIleLeuPheIle 245  
 Qy 458 CTTGCTATGGGATGGCGGTAGATGCAATGTTCTTGTATTTCGAAGAGATCCGAGAGAA 517  
 Db 246 LeuValIleGlyAlaAlaThrAspThrSerLeuLeuThrValAlaAlaArgPheArgGlu 265  
 Qy 518 TTTTATTGTCCTCAAGCTCTTAAAAAATCTGTAGAAAAAGGATATACCAAGCTTTTTCGA 577  
 Db 266 LeuArgValGlnGlnAspLysGlyIleAlaThrGlyLysAlaIleArgAlaSerValGlu 285  
 Qy 578 GCCATTTTGTGATCTTAACACTGACATGATATTGGCCTCAGCACTTCTTTCTCTCATAG 637  
 Db 286 ProIleLeuAlaSerGlySerThrValIleAlaGlyLeuLeuCysLeuPheSerAsp 305  
 Qy 638 -----ACAGGCGCTATTAAAGGTTTGTGTTGACATTGATTATTAGGA 679  
 Db 306 LeuLysSerAsnSerThrLeuGlyProValAlaSer-----ValGly 319  
 Qy 680 ATTTCCTCTTCAAGTTTACGGCTCTT---TTCATGACTAAATTTTCTTCATG----- 730  
 Db 320 IleIlePheAlaMetLeuSerAlaLeuThrLeuLeuProAlaLeuLeuPheValPheGly 339  
 Qy 730 ----- 730

Db 340 ArgValAlaPheTrpProlysaArgProlytyrGluProGluLysAlaArgAlaLysAsn 359  
 QY 731 -----CTGTGATGAATAGACCCACATACACAGTTGATGATG 772  
 Db 360 AspiLeProAlaSerGlyIleTirPserLysValAlaAspLeuValGluGlnHisProArg 379  
 QY 773 AATAAGTTCGTGGGATAAAGCATGATTTCTGAGAGAGATGCCAA----- 817  
 Db 380 AlaIleTrpValSerThrLeuIleValLeuLeuGlyAlaAlaPheValProThrLeu 399  
 QY 818 AAACCTTGGGCTGTTGGAGAGTGTTCCTTTAGTGGCTGCTCGGGTTTGA 877  
 Db 400 LysAlaAspGlyValSerGlnSerAspLeuValLeuGlySerSerGlnAlaArgAspGly 419  
 QY 878 GCCTGGAATTCGTTTGGAGATGATTTTAAAGAGAG-----TATGCC 922  
 Db 420 -----GlnGlnAlaLeuGlyGlnHisPheProGlyGlySerGlySerProAlaTyIle 437  
 QY 923 TTAACTTTAATCCAAAAGACATGCATGCATGCTTGTCTCAAAATGCGTGCAGAAAGTT 982  
 Db 438 IleValAspGluThrGlnAlaAlaGlnAlaAlaAspValVal----- 451  
 QY 983 GTGCATTAAGTACAGAGAGCGTCTTCTTCTAGAGACTTCGCTATTCAAACATTTGA 1042  
 Db 452 -----LeuAsnAsnAspAsnPheGluThrValThrValThr 463  
 QY 1043 TCTTCAAGAAAAGATCAAAATCTATTATTTAGATAGCTTAAAGCTTAACTAAGCAGATA 1102  
 Db 464 SerAlaAspSer-ProSerGlySerAlaProIleThrAlaAspGlyIleVal----- 480  
 QY 1103 CGAGCTCTCTCTCTAAATTAACATCATGACATGCGCGTTATTTGGAGTTGTTCAGA 1162  
 Db 481 -----ProLeuGly-----SerGlyThrAlaProGlyProValVal 493  
 QY 1163 AACAGCGCTGATTTCTCTACGAAACTCTA-----AACGAAAGCA 1204  
 Db 493 IGluGlyGlnValLeuLeuGlnAlaThrLeuValGluAlaProAspSerGluGlnAlaGly 513  
 QY 1205 AAATTTGTCAAGGTAGACAGCAAACTATCGAAGAAATGCGTTATCGAGCGACCATC 1264  
 Db 513 nLys-AlaIleArgSerIleArgGlnThrPheAlaAspGluAsnIleSer-AlaValVal 532  
 QY 1265 GGGCTTTAGAGCTTTGGCAATCATCTGCTCTATGTGAGTTGCCCTTTGAATGGCA 1324  
 Db 533 GlyGlyValThrAlaThrSerVal----- 540  
 QY 1325 TATGCTTCAGTGGCGGATGCGCTTAAATTCATGACCTT-----TTGGCT 1369  
 Db 541 -----AspThrAsnAspAlaSerIleHisAspArgAsnLeuIleIleProIleVal 557  
 QY 1370 ACCTGTCAGCTCTGTTTACGACATTTCTTTTGAAGAAATTCAAATGATTTGCA 1429  
 Db 558 LeuLeuValIleLeuValIleLeuMetLeuLeuLeuArgSerIleValAlaProLeuLeu 577  
 QY 1430 GCCATGCGTCTTAAATGATGTTATTTGGGGTATTCATTAACAATPACTTTCATTTT 1489  
 Db 578 LeuValValThrThrValValSerPheAlaThrAlaLeuGlyValAlaAlaLeuLeuPhe 597  
 QY 1490 GATGTAATTCGTAAGATCGACAGCGACGTTTACCCTATGACATGTTTATGTT--- 1546  
 Db 598 AsnIleValIlePheSerPheProGlyAlaAspProAlaValProLeuTyGlyPheValPhe 617  
 QY 1547 -----AATGATGCCCTTCAAAAAGAGCTTCAGCGGACGGTA 1582  
 Db 618 LeuValAlaLeuGlyIleAspTyraAsnIlePheLeuValThrArgIleArgGluGluThr 637  
 QY 1583 ATGACAAACAGCTACACTATACGTTTGTAAATGCTTTTGTATAGCGCGCTCTCT 1642  
 Db 638 LysThrHisGlyThrArgLeuGlyIleLeuArgGlyLeuThrValThrGlyGly----- 655  
 QY 1643 GTCTTAAATTTGGCATTTATATGACCATAGGAAATTCCTTA---GGAACCTTATTCGCT 1699

Db 656 -----ValIleThrSerAlaGlyValValLeuAlaAlaThrPheAlaAla 670  
 QY 1700 CTTTATATTGCACACCTCTCTGTTGTTTATGCTC 1735  
 Db 671 LeuTyValIleProIleLeuPheLeuAlaGlnIle 682  
 RESULT 44  
 AAB78979  
 ID AAB78979 standard; Protein; 730 AA.  
 AC AAB78979;  
 XX  
 XX  
 DT 30-APR-2001 (first entry)  
 XX  
 DE C. glutamicum SRT protein sequence SEQ ID NO:218.  
 XX  
 XX Corynebacterium glutamicum; stress; resistance; tolerance; SRT;  
 KW fine chemical production; organic acid; proteiogenic amino acid;  
 KW nonproteogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carboxylate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
 KW evolutionary study; environmental hazard; fermentation.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN MO200100804-A2.  
 PD  
 XX  
 PF 23-JUN-2000; 2000MO-IB00922.  
 XX  
 PR 25-JUN-1999; 99US-0141031.  
 PR 01-JUL-1999; 99DE-1030429.  
 PR 01-JUL-1999; 99US-0142692.  
 PR 08-JUL-1999; 99DE-1031413.  
 PR 08-JUL-1999; 99DE-1031457.  
 PR 08-JUL-1999; 99DE-1031541.  
 PR 09-JUL-1999; 99DE-1032209.  
 PR 09-JUL-1999; 99DE-1032230.  
 PR 14-JUL-1999; 99DE-1032914.  
 PR 27-AUG-1999; 99DE-1040764.  
 PR 27-AUG-1999; 99US-0151214.  
 PR 31-AUG-1999; 99DE-1041382.  
 XX  
 PA (BAD ) BASE AG.  
 XX  
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Lee H;  
 PI Kim H;  
 XX  
 DR WPI; 2001-061972/07.  
 DR N-PSDB; AAF71092.  
 XX  
 XX  
 PT New isolated Corynebacterium glutamicum nucleic acid encoding a stress,  
 PT tolerance or resistance protein, for production or modulation of  
 PT production of fine chemicals, such as, e.g. amino acids, lipids,  
 PT carbohydrates, or enzymes -  
 XX  
 PS Claim 20; Page 397-399; 526pp; English.  
 XX  
 CC AAF70984 to AAF71133 encode the Corynebacterium glutamicum stress,  
 CC resistance, and tolerance (SRT) proteins given in AAB78871 to AAB79020.  
 CC The C. glutamicum SRT genes (I) can be used in vectors (II) for  
 CC expression in host cells and production of fine chemicals, such as, an  
 CC organic acid, a proteiogenic or nonproteogenic amino acid (preferred),  
 CC a purine or pyrimidine base, a nucleoside, a nucleotide, a lipid, a  
 CC saturated or unsaturated fatty acid, a diol, a carboxylate, an aromatic  
 CC compound, a vitamin, a cofactor, a polyketide, or an enzyme. The fine  
 CC chemical production can be modulated. The presence of (I) or the SRT  
 CC proteins (III) encoded by them are used for diagnosing the presence  
 CC or activity of Corynebacterium diptheriae. (I), (II), (III) and host  
 CC cells containing them can be used to map the genomes of organisms related  
 CC to C. glutamicum, to identify and localise C. glutamicum sequences of  
 CC interest, in evolutionary studies, in determination of SRT protein

CC regions required for function, in modulating the SPT protein activity,  
 CC and in modulating the activity of an SKI pathway. (ii) are used to permit  
 CC C. glutamicum to survive in an environment that is normally  
 CC environmentally or chemically hazardous to it. (i) and protein molecules  
 CC encoded by it increase the survival of C. glutamicum to chemical and  
 CC environmental hazards and provide a means for continued growth and  
 CC multiplication in large scale fermentative growth conditions. By  
 CC increasing the growth rate or maintaining a normal growth rate in poor or  
 CC toxic conditions, the yield, production and/or efficiency or production  
 CC of fine chemicals from a culture may be increased.

XX Sequence 730 AA;

#### Alignment Scores:

Pred. No.:	3,6e-05	Length:	730
Score:	130.50	Matches:	128
Percent Similarity:	33.55%	Conservative:	78
Best Local Similarity:	20.85%	Mismatches:	270
Query Match:	3.94%	Indels:	138
DB:	22	Gaps:	23

US-09-868-987-1 (1-1864) x AAB78979 (1-730)

Qy	125	GTCCCATGAAAAATCATGCCAGTCTCTCAGGAAATTTACCACCGTGAAGTGAGCAA	184
Db	132	ValProLeuAsnProSerAlaGluLeuThr-----GluSerValGluLys	146
Qy	185	CTCGCCTCAGATTTAAATATCTCGGCGATGCTTTTGTTCGCGAG---GTTCTCAGTGAA	241
Db	147	LeuSerGluThrLeuThrGlnGlnThrProAspTyrValSerThrTyrValThrGlyPro	166
Qy	242	GAGACGATCTCTTCTGATCTTGGGAAAAACAATGT---ACACAGGCAATATCTCAGCA	298
Db	167	AlaGlyPheThrAlaAspLeuSerAlaAlaPheAlaGlyIleAspGlyLeuLeuAla	186
Qy	299	TGCTCGCTTGGCAATCTTATCTTTGATGAGCGTATATATAGATT---	349
Db	187	ValAlaLeuAlaValLeuValIleLeuValIleValTyrArgSerPheIleLeuPro	206
Qy	350	GGAGCGCTATCGCTTCGGGA-----GCTGTTCTCTGAATCTTTTCTTATCTGGCA	403
Db	207	IleAlaValLeuAlaThrSerLeuPheAlaLeuThrValAlaLeuLeuValTrpTrp	226
Qy	404	GCTTACAGTATTGGATCGCCACTACCTGTGTCAGGACTCGCTGGGATT-----GTT	457
Db	227	LeuAlaLysTrp---AspIleLeuLeuSerGlyGlnThrGlnGlyIleLeuPheIle	245
Qy	458	CTTGCTATGGGATGCGGTAGATCAATGTTCTTGTATTGGAAGATCCGAGAGAA	517
Db	246	LeuValIleGlyAlaAlaThrAspTyrSerLeuLeuTyrValAlaArgPheArgGluGlu	265
Qy	518	TTTTTATTGCTCAAGATCTTAAAAAATCTGAGAAAAAGGATATACCAAGCTTTTGA	577
Db	266	LeuArgValGlnAsnLysGlyIleAlaThrGlyLysAlaIleArgAlaSerValGlu	285
Qy	578	GCCATTTTGTATCTPAATGACTACAGTATTGGCCTCAGCACTTCTTTCTCTAGAT	637
Db	286	ProIleLeuAlaSerGlySerThrValIleAlaGlyLeuLeuCysLeuLeuPheSerAsp	305
Qy	638	-----ACAGGCGCTATTAAAGGTTTGTCTTCACATTGATTGATTAGGA	679
Db	306	LeuLysSerAsnSerThrLeuGlyProValAlaSer-----ValGly	319
Qy	680	ATTTTCTCTCAATCTTTACGCTCTT---TTCATGACTAAATTTTCTTCATG-----	730
Db	320	IleIlePheAlaMetLeuSerAlaLeuThrLeuLeuProAlaLeuLeuPheValPheGly	339
Qy	730	-----	730
Db	340	ArgValAlaPheTrpProLysArgProLysTyrGluProGluLysAlaArgAlaLysAsn	359
Qy	731	-----CTGTGGATGAATAGACCCCAACATACACATTCGATGATGATG	772

Db	360	AspIleProAlaSerGlyIleTrpSerLysValAlaAlaAspLeuValGluGlnHisProArg	379
Qy	773	AATAAGTTCTGGGATAAAGCATGATTCTTGTAGAGGATGCAAA-----	817
Db	380	AlaIleTrpValSerThrLeuIleValLeuLeuGlyAlaAlaPheValProThrLeu	399
Qy	818	AAACTTTTGGGCTGTTCTGAGAGTGTCTTTCTTTTGTAGGCTGCTCTCGGTTTGA	877
Db	400	LysAlaAspGlyValSerGlnSerAspLeuValLeuGlySerSerGluAlaArgAspGly	419
Qy	878	GCCTGGAATTCGTTTGGGAATGGATTTTAAAGGAGGG-----TATGCC	922
Db	420	-----GlnGlnAlaLeuGlyGluHisPheProGlySerGlySerProAlaTyrIle	437
Qy	923	TTTACCTTTAATCAAAAGAGCATGGCATCAGCGATGTTGCTCAAAATCGCTGGCAAGTT	982
Db	438	IleValAspGluThrGlnAlaAlaGlnAlaAlaAspValVal-----	451
Qy	983	GTGCATAAATACAGGAAGCTGCTTTCTTCTAGACACTTCGCTATTCAACATTTGGA	1042
Db	452	-----LeuAsnAsnAspAsnPheGluThrValThrValThr	463
Qy	1043	TCTTCAGAAAAAGATCAAAATCTATTTTAGTGATAAAGCTTTAAGCTATATAAGCAGATA	1102
Db	464	SerAlaAspSer-ProSerGlySerAlaProIleThrAlaAspGlyIleVal-----	480
Qy	1103	CGAGCCTCTCTCTCAAAATTAACGATCATGAGCTGGCGTTATTGTGGGATTTGTGAGA	1162
Db	481	-----ProLeuGly-----SerGlyThrAlaProGlyProValVal	493
Qy	1163	AACAGGCTAGATTCTCTACGGAAACTCTA-----AACGAAACGCA	1204
Db	493	lGluGlyGlnValLeuLeuGlnAlaThrLeuValGluAlaProAspSerGluGluAlaG	513
Qy	1205	AAATTTTGGTCAAGGTAAAGCAGCAAACTATCGAAGAAAATTCGCTTATCAGGCCACATC	1264
Db	513	nLys-AlaIleArgSerIleArgGlnThrPheAlaAspGluAsnIleSer-AlaVal	532
Qy	1265	GGGCTTTTAGAGCTTTGGCAATCATCTGCTCTATGAGTTTTCGCTTTGAATGGCA	1324
Db	533	GlyGlyValThrAlaThrSerVal-----	540
Qy	1325	TATGCTTTTCAAGTCCGATGCGCTTTAATTCATGACCTT-----TTGGCT	1369
Db	541	-----AspThrAsnAspAlaSerIleHisAspArgAsnLeuIleIleProIleVal	557
Qy	1370	ACCTGTGAGCTCTGTTTATAGACACATTTCTTTTGAAGAAAATTCAAATAGATTGCAA	1429
Db	558	LeuLeuValIleLeuValIleLeuMetLeuLeuLeuArgSerIleValAlaProLeuLeu	577
Qy	1430	GCCATTGCTGCTTTAATGACTGTTATGGGGTATTTCATTAACAATACTTTTGATCATTTT	1489
Db	578	LeuValValThrThrValValSerPheAlaThrAlaLeuGlyValAlaAlaLeuLeuPhe	597
Qy	1490	GATCGTATTCTGTAAGATCGCCAAAGCAACCTGTTTACCCTATGCTATGTTTATGTT---	1546
Db	598	AsnHisValPheSerPheProGlyAlaAspProAlaValProLeuTyrGlyPheValPhe	617
Qy	1547	-----AATGATGCCCTTCAAAAGACGTTTCAGCCGACCGTA	1582
Db	618	LeuValAlaLeuGlyIleAspTyrAsnIlePheLeuValThrArgIleArgGluThr	637
Qy	1583	ATGACAACAGCTACAACTCTATCAGTTTGTAAATGCTTTTGTATAGGGCGCTCTCT	1642
Db	638	LysThrHisGlyThrArgLeuGlyIleLeuArgGlyLeuThrValThrGlyGly-----	655
Qy	1643	GTCTTTAATTTTGCATTTATATGACCATAGGATCTTCTTA---GGAACCTTTATCGCT	1699
Db	656	-----ValIleThrSerAlaGlyValValLeuAlaAlaThrPheAlaAla	670
Qy	1700	CTTTATATTGACCACTCTGTTGTTTATGTC	1735
Db	671	LeuTyrValIlePheLeuPheLeuAlaGlnIle	682







Db 690 -----TrpValProGlnLeuMetGluLysLeuGlnGlnLe 701  
Qy 1435 TGGTCTTTAATGACTGTATTGGGTATTCAATAACAATACTTTTGATCATTTTGTGATCG 1494  
: ||| ||| : : : : : : : : : : : :  
Db 701 uProGlnLeuSerAspValSerSerAspTyrGlnAspLysGlyLeuValAlaTyrValAs 721  
: : : : : : : : : : : : : : : : : :  
Qy 1495 TATTGGTGAAGATCGCCCAAGCGAACCTGTTTACCCCTATGCATGTTTGTAGTTAATGATGC 1554  
: : : : : : : : : : : : : : : : : :  
Db 721 nValAspArgAspSerAlaSerArgLeuGlyIleSerMet---AlaAspValAspAsnAl 740  
: : : : : : : : : : : : : : : : : :  
Qy 1555 CCTTCAAAAAGACGTTTCAGCCGCACGGTAATGACACACAGCTACAACT 1600  
||| ||| : : : : : : : : : : : :  
Db 740 aLeuTyraaAlaPheGlyGlnArgLeuIleSerThrIleTyrThr 755

Search completed: December 17, 2002, 12:35:33  
Job time : 103.5 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 17, 2002, 12:30:02 ; Search time 57.5 Seconds  
(without alignments)  
13359.020 Million cell updates/sec

Title: US-09-868-987-1  
Perfect score: 3312  
Sequence: 1 agggacttcgcatattgtc.....atgacaaattcgataatgc 1864

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlh  
-DB=SPTRMBL 21 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=45 -MODE=LOCAL  
-OUTFMT=ptc -NOR=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US09868987 @CGN 1 1 70 @runat 16122002 083006 23540 -NCPU=6 -ICPU=3  
-NO XLPXY -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WEN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6 -FGAEXT=7  
-YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2874	86.8	1402	16 Q927Y9	Q927Y9 chlamydia p

2	2083	62.9	1400	16	084454
3	2088	62.4	1400	16	084454
4	840	25.4	876	16	084454
5	790.5	23.9	849	16	084454
6	715.5	21.6	845	16	084454
7	685	20.7	759	16	084454
8	682	20.6	858	16	084454
9	682.5	20.6	844	16	084454
10	678.5	20.5	737	16	084454
11	668.5	20.2	768	16	084454
12	653.5	19.7	775	16	084454
13	630	19.0	795	16	084454
14	620.5	18.7	754	16	084454
15	618.5	18.7	754	16	084454
16	522.5	15.8	873	16	084454
17	460.5	13.9	411	16	084454
18	436	13.2	474	16	084454
19	424	12.8	620	16	084454
20	423.5	12.8	419	16	084454
21	403.5	12.2	526	16	084454
22	400.5	12.1	518	2	084454
23	399.5	12.1	616	16	084454
24	395.5	11.9	518	16	084454
25	391	11.8	417	16	084454
26	391	11.8	532	16	084454
27	390	11.8	323	16	084454
28	384.5	11.6	322	16	084454
29	383.5	11.6	615	16	084454
30	381.5	11.5	410	16	084454
31	376	11.4	617	16	084454
32	373	11.3	626	16	084454
33	369.5	11.2	306	16	084454
34	368	11.1	612	16	084454
35	365.5	11.0	437	16	084454
36	365	11.0	323	16	084454
37	364	11.0	618	16	084454
38	364	11.0	618	16	084454
39	363.5	11.0	315	16	084454
40	357.5	10.8	614	16	084454
41	357.5	10.8	615	2	084454
42	356.5	10.8	615	16	084454
43	356.5	10.8	615	16	084454
44	356	10.7	615	2	084454
45	355.5	10.7	465	16	084454

## ALIGNMENTS

### RESULT 1

ID	Q927Y9	PRELIMINARY;	PRT;	1402	AA.
AC	Q927Y9;				
DT	01-MAY-1999	(TRENBLrel. 10, Created)			
DT	01-MAY-1999	(TRENBLrel. 10, Last sequence update)			
DT	01-JUN-2002	(TRENBLrel. 21, Last annotation update)			
DE	Protein export proteins SECDF/SECF (Fusion) (SECDF protein, putative).				
DE	SECDF/SECF OR SECDF/SECF OR CPN0564 OR CP0186.				
GN	Chlamydia pneumoniae (Chlamydia pneumoniae).				
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).				
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.				
OX	NCBI_TaxID=83558;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CW029;				
RX	MEDLINE=99206606; PubMed=10192388;				
RA	Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W., Olinger L., Grimwood J., Davis R.W., Stephens R.S.;				
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";				
RL	Nat. Genet. 21:385-389(1999).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=AR39;				
RX	MEDLINE=20150255; PubMed=10684935;				



Qy	1559	CAAAAGACGTTACAGCGCACCGGTAATGACAAACAGCTACAACTCTATCATGTTTGTAAATG	1618
Db	1337	GlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMet	1356
Qy	1619	CTTTTGTATTAGCGCGCTCTCTGCTTTTAATTTTGCATTTATTATCATCACCATAGGGATT	1678
Db	1357	LeuLeuPheIleGlyGlySerValPheAsnPheAlaPheIleMetThrIleGlyLe	1376
Qy	1679	CTTCTAGGAACTTTATCGTCTCTTTATATATGCACCACCTCTGTTCTTTTATGTCGTCGT	1738
Db	1377	LeuLeuGlyThrLeuSerSerLeuTyrlleAlaProProLeuLeuLeuPheMetValArg	1396
Qy	1739	AAAGAAAATCGCTCAAAA	1756
Db	1397	LysGluAsnArgSerLys	1402
RESULT 2			
ID	084454	PRELIMINARY; PRT; 1400 AA.	
AC	084454;		
DT	01-NOV-1998 (TrEMBLrel. 08, Created)		
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Protein export.		
GN	SECD/SECF OR CT448.		
OS	Chlamydia trachomatis.		
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.		
NCBI_TaxID=813;	[1]		
SEQUENCE FROM N.A.			
RC	STRAIN=D/UW-3/CX;		
RX	MEDLINE=99000809; PubMed=9784136;		
RA	Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,		
RA	Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,		
RA	Davis R.W.;		
RT	Genome sequence of an obligate intracellular pathogen of humans:		
RT	Chlamydia trachomatis."		
RL	Science 282:754-759(1998).		
DR	EMBL; AB001318; AAC68047.1; -.		
DR	InterPro; IPR001969; Asparticase site.		
DR	InterPro; IPR001005; Myb_DNA_binding.		
DR	InterPro; IPR003335; SecD_SecF.		
DR	Pfam; PF02355; SecD_SecF; 1.		
DR	TIGRFAMS; TIGR00916; 2A0604s01; 2.		
DR	TIGRFAMS; TIGR01129; secD; 1.		
DR	PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.		
DR	PROSITE; PS00037; MYB_1; UNKNOWN_1.		
KW	Complete proteome.		
SQ	SEQUENCE 1400 AA; 156514 MW; 09400F968AD18406 CRC64;		
Alignment Scores:			
Pred. No.:	5,79e-168	Length:	1400
Score:	2083.00	Matches:	415
Percent Similarity:	79.28%	Conservative:	67
Best Local Similarity:	68.26%	Mismatches:	80
Query Match:	62.89%	Indels:	46
DB:	16	Gaps:	5
US-09-868-987-1 (1-1864) x 084454 (1-1400)			
Qy	(2	TGACATTCCGATATTGTCCAGGAGGGATCAGCGCAGCTGCTAATGACAAATTTCTGCA	61
Db	812	TpThrSerLysPheCysGlnGluGlyValSerGlyThrLysAsnGlyLeuPheSerGly	831
Qy	62	AACGCTGGATGGGTATGGCTGTAGTGTATGACGGTTATATGCTCAGACGCCCTATTTTA	121
Db	832	GlyArgGlyTrpArgMetAlaValLeuApsGlyTyrValIleSerAspProValLeu	851
Qy	122	AACGTCCCATTTGAAATCATCGCTGCTCAGGGMAATTTACCCACCGTGAAGTGAGC	181
Db	852	AsnValProLeuLysAspHisAlaSerValSerGlyAsnPheSerTyrArgLysValHis	871
Qy	182	AAACTCGCCTCAGATTAAATCTGAGCGCATGCTCTTTGTTCCCGAGTCTTCAGTGAA	241

872	ArgLeuAlaAlaAspLeuIysSerGlyAlaMetSerPheIleProGluIleLeuSerGlu	891
242	GAGACGATCTCTTCATCTCTGGGAAAAAACAATCTACACAAGGCATTATCTCAGCATGC	301
892	GluValValSerProGluLeuGlySerSerGlnArgValGlnGlyIleLeuSerValVal	911
302	TGTGGCTTGGCAATGCTTATTGTTTGTATGACGCTATATTATAGATTTGGAGCGTCATC	361
912	LeuGlyLeuValValLeuIleValLeuMetSerValTyrTyrArgPheGlyGlyValIle	931
362	GCTTGGGAGCTGTTCTCTCGAATCTTTTGTCTATCTGGGAGCTCTACAGTATTTCGAT	421
932	AlaSerIleAlaValLeuLeuAsnLeuLeuLeuIleTyrAlaSerMetGlnTyrLeuAsp	951
422	GCGCCATCACTTGTTCAGGACTCGCTGGGATTGTTCTTGTATGGGATGCCGCTAGAT	481
952	AlaProLeuThrLeuSerGlyLeuAlaGlyIleIleLeuAlaMetGlyMetAlaValAsp	971
482	GCAATGTTCTGTATTCGAAGAATCCGAGAGGAATTTTATTGTCTCAAAGTCTTAAA	541
972	AlaAsnValLeuValPheGluArgIleArgGluGluTyrLeuLeuThrArgSerLeuSer	991
542	AAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGGCATTTTGTATTCTTAACCTGACT	601
992	GluSerValGluAlaGlyTyrLysLysAlaPheSerAlaIlePheAspSerAsnLeuThr	1011
602	ACAGTATTGGCCTCACGACTCTCTTCTCTCTAGATACAGGGCCTATTAAAGGGTTTGCT	661
1012	ThrIleLeuAlaSerAlaLeuLeuLeuMetLeuAspThrGlyPheIleLeuGlyPheAla	1031
662	TTGACATTTGATTTTAGGAATTTTCTCTCAATGTTTACGGCTCTTTTCATGACATAAATTT	721
1032	LeuThrLeuIleIleGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLysPhe	1051
722	TTCTTCAATGCTGTGATGAATTAAGACCCAAACATACACAGTTGCATATCATGAATAAGTTC	781
1052	PhePheValIleTyrValGlnLysThrArgGluThrGlnLeuHisMetMetAsnLysPhe	1071
782	GTGGGATAAAGCATCATTTCTTGAGAGGATCCAAAAAATTTTGGCGTGTCTTCTGGAAGT	841
1072	IleGlyIleLysHisAsnPheLeuLysGlyLysLysArgLeuTyrValValSerGlyVal	1091
842	GTTTCTTTTAGTGTGCGTGTCTCTCGGGTTTGGAGCTCGAATTCGGTTCGGGAATG	901
1092	ValLeuValLeuGlyCysValGlyLeuGlyPheGlyAlaTyrAspSerValPheGlyMet	1111
902	GATTTAAAGAGGGTATGCTTTACCTTTAATCCAAAGAGCATGGCATCAGCGATGTT	961
1112	AspPheLysGlyTyrAlaLeuThrLeuAsp-----SerAspVal	1125
962	GCT-----CAAATGCGTGCAGGTTGTGCATAACTACAGGAAGCTGGT	1006
1126	CysGluTyrAsnProGluGlnMetCysSerValLeuArgLysArgPheGlnGlnIleGly	1145
1007	CTTCTCTCTAGACACTTCGGTATTCAACATTTGCATCTTCAGAAAAAGATCAAAATCTAT	1066
1146	LeuSerSerArgAspTyrArgValArgAlaAspSerSerGluLysValLysIleTyr	1165
1067	TTTAGTGATAAGCTTTAAGCTATACTAAGCAGATACGA-----	1105
1166	LeuSerGlnAsnAlaLeuAspArgValGluGlnIleGluGlyAlaGlySerGluGlnLys	1185
1106	-----GCCCTCTCTCTTAAATTAACGATCATGAGCTGGCGTTATTGTT	1147
1186	GlySerAspTyrHisLeuAlaArgValLeuGlnValLeuSerAspSer-----	1201
1148	GGGATTGTTGTAGAAACAGGCGCTAGATTCTCTACGGAAACTCTTAAACGAAACGCAAAA	1207
1202	-----GlySerSerThrThrSerMetVal	1209
1208	TTT-----TGGTCAAAGGTAAAGCAGCAAACTATTCGAGAAAAATCGGT	1249

Db 1210 PheAspAlaSerArgGlySerTyrPheLeuValSerGlyGlnLeuSerAsnLysMetArg 1229  
 Qy 1250 TATCAGCGCAGCATCGGGCTTTTACGAGCTTTGGCATCATCTGCTTATGTGACTTTG 1309  
 Db 1230 ThrGlnAlaValIleAlaLeuPheGlyAlaLeuGlyIleIleLeuPheTyrValSerLeu 1249  
 Qy 1310 CGCTTTGAATGGCAATATGCTTTCAGTCGCGATGCGGATGCGGTTTAAATGACTTTGGCT 1349  
 Db 1250 ArgPheGlnTyrPargTyrAlaPheSerAlaIleCysAlaLeuMetHisAspLeuLeuAla 1269  
 Qy 1370 ACCGTGCGACTCTGTTTATATGACATTTCTTTTGAAGAAATTCAAATGATTTGGCA 1429  
 Db 1270 ThrCysAlaValIleLeuValAlaLeuHisPhePheLeuGlnArgIleGlnIleAspLeuGln 1289  
 Qy 1430 GCCATGTGCTGCTTATATCATCTGATTTGGGATTCATTAACAATACCTTGTATCTTTT 1489  
 Db 1290 AlaIleGlyAlaLeuMetThrValLeuGlyIleSerLeuLeuAsnThrLeuIleIlePhe 1309  
 Qy 1490 GATGCTATTCGTGAAGATCGCGAAGCGAACCTGTTTACCCCTATGCGATGTTTGGTTAT 1549  
 Db 1310 AspArgIleLeuValGlnAspArgGlyIleValLeuPheMetProIleLeuLeuLeuAsn 1329  
 Qy 1550 GATCCCTCTTCAAAAGACCTTCAGCGCGACGCGTAATGACACAGCTACACTCTATCAGTT 1609  
 Db 1330 AspAlaLeuGlnIleValThrLeuGlyArgThrValMetThrThrAlaThrThrLeuSerVal 1349  
 Qy 1610 TTGTTAATGCTTTTGTATTAAGCGCGCTCTCTGCTTTAAATTTTGCAATTTATATGACC 1669  
 Db 1350 LeuValIleLeuLeuPheValGlyIleGlyIleSerIlePheAsnPheAlaPheIleThr 1369  
 Qy 1670 ATAGGAGATCTCTTACGAACTTTTACGCTCTTATATATGACACACCTGTTGTTGTTT 1729  
 Db 1370 ValGlyIleLeuLeuGlyIleThrLeuSerSerLeuIleAlaProIleLeuLeuPhe 1389  
 Qy 1730 ATGCTCCCTAAAGAAATC3CTCA 1753  
 Db 1390 MetValArgLysGlnGlnGlnAsn 1397

## RESULT 3

Q9PJU4 PRELIMINARY; PRT; 1400 AA.  
 ID Q9PJU4  
 AC Q9PJU4  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE SecDF protein, putative.  
 GN TC0733.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83560;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MOBN / NIGG;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gail S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utechtack T., Betty K., Baas S.,  
 RA Linher K., Weidman J., Khoult H., Craven B., Bowman R.,  
 RA Gelin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia  
 RT pneumoniae A39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL; AE002341; AAF39543.1; -.  
 DR TIGR; TC0733; -.  
 DR InterPro; IPR001036; Acrflavin res.  
 DR InterPro; IPR001005; Myb DNA binding.  
 DR InterPro; IPR003335; SecD\_SecE.  
 DR Pfam; PF02355; SecD\_SecE; 1.  
 DR PRINTS; PR00702; ACRIFLAVINRP.  
 DR TIGRFAMS; TIGR00916; 2A0604s01; 2.  
 DR TIGRFAMS; TIGR01129; secD; 1.  
 DR PROSITE; PS00037; MYB\_1; UNKNOWN\_1.  
 KW Complete proteome.

SQ SEQUENCE 1400 AA; 156140 MW; 559E509CD823823F CRC64;  
 Alignment Scores:  
 Pred. No.: 1,08e-166 Length: 1400  
 Score: 2068.00 Matches: 407  
 Percent Similarity: 79.61% Conservative: 77  
 Best Local Similarity: 66.94% Mismatches: 78  
 Query Match: 62.44% Indels: 46  
 DB: 16 Gaps: 5

US-09-868-987-1 (1-1864) x Q9PJU4 (1-1400)

Qy 2 TGAGCTCCGCAATATGTCAGAGGGATCAGCGGACCTGTAATGCAATATTCGCA 61  
 Db 812 TrrThSerLysPheCysGlnGlnGlyValAsnGlyThrLysAsnSerGlnPheSerGly 831  
 Qy 62 AACCTGATGAGCGGTATGCGTATGATGATGACGTTATATGCTACAGACCTTATTTTA 121  
 Db 832 GlyArgGlyTrrPargMetAlaValValLeuAspGlyTyrValIleSerAspProValLeu 851  
 Qy 122 AACGTCCCATGAAATATCATGCGCACTGCTTCAGGAAATTTACCAACCGTGAATGAGC 181  
 Db 852 AsnValProLeuLysAspHisAlaSerValSerGlyAsnPheSerTyrArgGlnValAsn 871  
 Qy 182 AACCTGCGCTCAGATTTAAATCTGGAGCGATGCTTTTGTCCGAGGCTTCAGTGA 241  
 Db 872 ArgLeuAlaSerPheLysSerGlyAlaMetSerPheIleProGlnIleLeuSerGln 891  
 Qy 242 GAGACGATCTCTGTATCTTGGGAAAAAACAATGTACACACAGCAATATCTCAGATGC 301  
 Db 892 GlnValValSerProGlnLeuGlyHisSerGlnArgLeuGlnGlyIleLeuSerValPhe 911  
 Qy 302 TGGGCTTGGCAATGCTTATGTTTGTATGAGCGGTATATATGATTTGGAGCGGTATC 361  
 Db 912 LeuGlyLeuValValLeuIleAlaLeuMetSerValTyrTyrLysPheGlyValIle 931  
 Qy 362 GCTTGGGAGCTGTTCTGTGAATCTTTGCTATCTGAGCGACCTCAGATATTTGGAT 421  
 Db 932 AlaSerValAlaValLeuLeuAsnLeuLeuIleTrrPAlaSerMetGlnTyrLeuAsp 951  
 Qy 422 GCGCACTCATCTTGTTCAGACTCGCTGGGATGTTCTTGTGATGAGGATGCGGTAGAT 481  
 Db 952 AlaproLeuThrLeuSerGlyLeuAlaGlyIleIleIleLeuAlaMetGlyMetAlaValAsp 971  
 Qy 482 GCAATGTTCTTGTATTCGAAAGATCCGAGAGAAATTTTATGTTCTCAAGCTTTAA 541  
 Db 972 AlaAsnValLeuValPheGlnArgIleArgGlnGlyIleLeuLeuSerArgSerLeuSer 991  
 Qy 542 AAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCAATTTTGTATGATCTTACTGACT 601  
 Db 992 GlnSerValGlnAlaGlyTyrLysValaPheSerAlaIlePheAspSerAsnLeuThr 1011  
 Qy 602 ACAATATGCGCTCAGACATCTTCTTCTTCTAGATACAGGCGCTATTAAGGATTTGCT 661  
 Db 1012 ThrIleLeuAlaSerAlaLeuLeuLeuMetLeuAspThrGlyProIleLysGlnPheAla 1031  
 Qy 662 TTGACATGATTTTAGAAATTTTCTCTTCATATGCTTACGCTCTTTCATGACTTAATTT 721  
 Db 1032 LeuThrLeuIleIleGlyIlePheSerSerMetPheThrSerLeuPheMetThrLysPhe 1051  
 Qy 722 TTCTCATCGTGGATGAATGAAGCAATACCAATACATACAGTGTGATATGAGATTAAGT 781  
 Db 1052 PhePheLeuIleTrrPAlaGlnLysTrnArgGlnThrGlnLeuHisPheMetCysLeuLysPhe 1071  
 Qy 782 TGAGGATTAAGCATGATTTCTTGAGAGATGCAAAAACCTTTGGGCTGTTCTGAGAT 841  
 Db 1072 IleGlyValLysHisAsnPheLeuLysGlnCysLysArgLeuTrrPAlaValSerGlyAla 1091  
 Qy 842 GTTTTCTTTTAAAGTTCGTTGCTCTCGGGTTTGGAGCTCGAATTCGCTTTTGGGATG 901  
 Db 1092 ValValValLeuGlyCysIleGlyLeuGlyPheGlyAlaTrrPAspSerIleLeuGlyMet 1111  
 Qy 902 GATTTTAAAGAGGATATGCTTACTTAAATCCAAAAGACATGCGCATCAGCATGTT 961









## Alignment Scores:

Pred. No.:	4,086-52	Length:	845
Score:	715.50	Matches:	175
Percent Similarity:	52.87%	Conservative:	129
Best Local Similarity:	30.43%	Mismatches:	198
Query Match:	21.60%	Indels:	73
	16	Gaps:	14

US-09-868-987-1 (1-1864) x Q8YH6 (1-845)

```

QY 80 GCTGTAGTATGACGGTTATATGTCACAGCCCTATTATTAACCTCCATGAAA--- 136
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 298 AlalleValleuaspGlyValleuSerAlaProValIleArgGluProIleThrGly 317
QY 137 AATATGCGCAGTCTCTGAGGAAATTTACCACTGATGAGCAAACTGCCTCAGAT 196
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 318 GlySerGlyGlnIleSerGlySerPheThrValGlnAspThrValIleLeuSerAlaLeu 337
QY 197 TTAATAATCTGAGCATGTCCTTTGTTCCGAGGTCCTCACTGAAGAGACGATCTTCT 256
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 338 LeuArgAlaGlyAlaLeuProAlaProLeuThrValIleGluGluArgThrValGlyPro 357
QY 257 GATCTGGGAAAAAACAATGTACACAGCATTAATCTCAGCATGCTGTGCTGGCAATG 316
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 358 AspleuGlyGlyAspAlaIleLysMetGlyLeuMetThrGlyIleIleGlyPheLeuLeu 377
QY 317 CTATGTTGTTGATGACGCTATATATAGATTGGAGCGCTCATGCTTCGGAGCTGTT 376
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 378 ValAlaValPheIleLeuLeuLeuIleuIleGlyPheTrpGlyValIleAlaAsnValAlaLeu 397
QY 377 CTTCGATATCTTTGCTTATCTGAGGAGCTCTACAGTATTTGATGAGCCACCTGCTG 436
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 398 LeuLeuHisThrLeuLeuThrPheSerAlaLeuSerLeuIleGlyAlaThrLeuThrLeu 417
QY 437 TCAGAGCTCGCTGGATGTTCTTCTATGCGGATGCGCTGATGCAATGTTCTGTA 496
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 418 ProGlyIleAlaGlyIleIleLeuGlyIleGlyIleAlaValAlaAsnIleLeuIle 437
QY 497 TTCGAAAGATCCGAGAGAAATTTTATGTCCTCAAGCTTTAAAAATCTGTGAAAAA 556
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 438 AsnGluArgIleArgGluGluThrArgLysGlyLeuGlyAlaMetAlaIleLeuAspLys 457
QY 557 GGAATATCCAGAGCTTTTGAGCATTTTGTATCTTCACTGATGATGAGCCCA 616
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 458 GlyPheHisSerAlaPheAlaThrIleValAspAlaAsnValIleThrLeuThrAlaThr 477
QY 617 GCACCTTCTTCTCTCTATACAGAGGCTATTAAGAGGTTGCTTGACATGATTTTA 676
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 478 IleLeuLeuPheLeuPheGlyThrGlyProValArgGlyPheAlaValIleThrMetMetLeu 497
QY 677 GGAATTTCTCTCAATGTTTACGCTCTTTTCATGACTAAATTTTCTTCATGCTGTG 736
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 498 GlyIleAlaIleSerMetPheThrAspValIleLeuValArg---MetIleMetAlaTrp 516
QY 737 ATGAATAG-----ACC 748
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 517 PheValAlaArgArgArgLysLeuValIleLeuHisIleGluProPheLeuLysPheValPro 536
QY 749 CAACATACACAGTGCATATATGTAAT---AAGTTCTGCGGATTAAGACATGATTTCTTG 805
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 537 GluHisThrIlePhePheArgPheMetCysAlaArgPheIleGlyIle----- 551
QY 806 AGAGGATGCAAAAACTTT36GCTGTTCTCGAAAGTGTCTTTTNAAGTTGCGGTGCT 865
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 552 ---GlyValSerIleValLeuSerIleAla---SerIleIleLeuPhePheLysProGly 569
QY 866 CTCGGGTTTGAGCGCTGGAATTCGCTTTTGGGAATGATTTTAAGAGGGTATGCTTCT 925
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 570 LeuAsnLys-----GlyIleAspPheLysGlyGlyIleGlnAla 582
QY 926 ACCTTATATCCAAAAGACATGATGATGATGATGATGATGATGATGATGATGATGATG 985
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 583 GluIleSerThrSerGlnPro-----AlaAspLeuAlaGlnIleuArgAlaLysLeu--- 599

```

```

QY 986 CATTAACATAGAGAGAGTGTCTTTCTCTAGAGACTCCGATTTCAACATTGGATCT 1045
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 600 -----GlyAlaLeuAsnLeuGlyValAlaIleuGlnThrIleArgLys 614
QY 1046 TCAGAAAGATCAAAATCTATTTT-----AGTATTAAGCTTTAAGC 1087
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 615 ProAsnGlnValLeuIleArgValGlnArgGlnGlyGlyGluGluAlaGlnThrAla 634
QY 1088 TATATCAAGCAATACAGAGCTCTCTCTAAATTAACGATCAATGAGCTGGCGTTATGCT 1147
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 635 AlaIleAsnLysMetArgGluAlaValIleThrGluLeuAsp-----Pro 648
QY 1148 GGGATTTGCTCAAAAACAGCGCTAGATTTCTTACGAAACTCTAAGCAAAACGAAAA 1207
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 649 GlyValIle-----GluArgThrGluVal 657
QY 1208 TTTTGGTCAAAGGTAAGCAACAACTATCAAGAAATGCGTTATCAGCGACCATCGGG 1267
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 658 ValGlyProLysValSerGlyGluLeuAlaArg-----SerGlyMetIleAla 673
QY 1268 CTTTAAAGACCTTTGGCAATCATCTTCTCTATGATGTTGGCTTGAATGCAATAT 1327
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 674 ValIleLeuSerAlaValAlaMetLeuPheIleTrpTrpArgPheGluTrpPhe 693
QY 1328 GCTTACAGTCCGCTATGCGCTTTAATTCATGACCTTTTGCTTACCTGTGCACTTGT 1387
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 694 AlaLeuGlyAlaIleAlaThrLeuIleLeuAsp-----ThrThrLysIle 708
QY 1388 ATAGACATATCTTTTGAAGAAATTCAAATGATTTGCAAGCCATGCTGCTTTAATG 1447
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 709 IleGlyPhePheAlaLeuThrGlnLeuAspPheAsnLeuThrAlaIleAlaLeuLeu 728
QY 1448 ACTGATTTGGGGTATTCATTAACAATCTTGATGATTTTGATGCTGGAAGAT 1507
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 729 ThrIleIleGlySerValIleAsnAspLysValValIleArgLysPheArgGluAsn 748
QY 1508 CGCCAGCGAAGCTGTTTACCTATGATGATGATGATGATGATGATGATGATGATG 1567
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 749 MetArgLeuThrLysSerLysThrLeuArgGluIleIleAspMetSerIleAsnGlnVal 768
QY 1568 TTCAGCGGACAGGTAATGACAAACAGTCACTCACTGATTTGTAATGCTTTGTT 1627
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 769 LeuValAlaGlySerIleThrThrSerMetThrThrPheLeuCysMetPheProMetAlaIle 788
QY 1628 ATAGAGGCTCTCTCTCTTAAATTTTGAATTTATGACATTAAGCAATGAGATTTCTTAGA 1687
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 789 TrpGlyGlySerAlaValHisAsnPheAlaValProMetLeuPheGlyIleValIleAla 808
QY 1688 ACTTATGCTCTTTATATGACCACTCTGTTGTTATG 1732
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 809 ThrSerSerSerIlePheIleAlaIleProIleLeuLeuLeu 823

```

## RESULT 7

099TL6 PRELIMINARY; PRT; 759 AA.

AC 099TL6;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Protein-export membrane protein SecDF.  
 GN SECDF OR SAVA1637 OR SAI1463.  
 OS Staphylococcus aureus (strain Mu50) / ATCC 700699), and  
 OC Staphylococcus aureus (strain N315).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Staphylococcus.  
 OX NCBI\_TaxID=158878, 158879;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);  
 RX MEDLINE=2131952; PubMed=1148146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,





```

Db 763 LeuserValleuileAspThrSerIleasnGlnThrLeuSerArgThrValLeuThrAla 782
QY 1592 GCTAACTCTATCAGTTTGTGTTAACTGTTGTTATAGCGGCTCTCT--GTCTTT 1648
Db 783 ValThrThrLeuLeuAlaLeuValAlaLeuThrLeuPheGlyGlySerGlyLeuValGln 802
QY 1649 AATTTTCATTTATATGACCATAGGATCTCTCTAGGAACCTTTATCGTCTCTTATTT 1708
Db 803 SerPheSerAlaThrLeuilePheGlyValLeuValGlyThrValSerSerIleThrVal 822
QY 1709 GCACCACTCTGTTGTTGTTATGCTCGTAAAGAAATGCG 1750
Db 823 AlaGlyProLeuLeuileLeuPheAsnArgArgAsnArg 836

RESULT 9
Q92S58 PRELIMINARY; PRT; 844 AA.
AC Q92S58
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Putative protein-export membrane protein.
GN SEC22 OR R00569 OR SMC02265.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021.
RX MEDLINE=2136507; PubMed=11481430;
RA Capela D., Barloy-Hubier F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Maguy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rameperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).
DR EMBL; AL591784; CAC45141.1; -.
DR InterPro; IPR001335; SecD SecF.
DR Pfam; PF02355; SecD_SecF; 1.
DR TIGRFAMs; TIGR00916; 2A0604s01; 2.
DR TIGRFAMs; TIGR01129; secD; 1.
KW Complete proteome.
SQ SEQUENCE 844 AA; 90315 MW; C313A0597E39D903 CRC64;

Alignment Scores:
Pred. No.: 2.55e-49 Length: 844
Score: 682.50 Matches: 172
Percent Similarity: 50.51% Conservative: 125
Best Local Similarity: 29.25% Mismatches: 208
Query Match: 20.61% Indels: 83
DB: 16 Gaps: 12

US-09-868-987-1 (1-1864) x Q92S58 (1-844)
QY 62 AACCGTGATGGCGTATGGTGTAGTGATTCAGCGTTATATGTCAGCGCCCTATTTTA 121
Db 291 AsnValGlyArgProPheAlaileValleuileAspGlyLysValLeuThralaProValle 310
QY 122 AACGTCCCA---TTGAAAAATCATGCCAGTGTCTCAGGGAATTTTACCACCGTGAAGTG 178
Db 311 AsnGluProIleLeuGlyGlyArgGlyGlnIleSerGlySerPheThrValGluGluAla 330
QY 179 AGCAACTCGCCTCAGATTAAATCTGAGCGGATGCTTTTGTTCGCGAGGTTCTCAGT 238
Db 331 ThrValLeuSerAlaLeuLeuArgSerGlyAlaLeuProAlaProLeuThrIleleu 350
QY 239 GAAGAGACGATCTCTCTGATCTTGGGAAAAACAATGTCACAGGCATTTATCTCAGCA 298
Db 351 GluArgSerValGlyProAsnLeuGlySerAspSerIleArgMetGlyLeuPheThrGly 370

```

```

QY 299 TGCTGTGGCTTGGCAATGCTTATTTGTTTATGAGCGCTATATATATAGATTGGAGGCGTC 358
Db 371 LeuAlaGlyPheGlyLeuValValValLeuMetValValLeuTyrGlyAlaTrpGlyMet 390
QY 359 ATCGCTTCGGGAGCTGCTTCTCTGAATCTTTTCTTATCTGGCAGCTCTACAGTATTG 418
Db 391 IleAlaAsnValGlyLeuValLeuHisThrIleMetThrIleGlyValLeuGlyLeuile 410
QY 419 GATGCGCCACTCACCTTGTCTGAGACTCGCTGGGATTGTTCTTGTATGGGATGGCGGTA 478
Db 411 GlySerThrLeuThrLeuProGlyIleAlaGlyIleleuLeuGlyIleGlyMetAlaVal 430
QY 479 GATGCAAAATGCTTGTGATTCGAAAGAAATCCGAGAGAATTTTATTGTCTCAAGTCTT 538
Db 431 AspAlaAsnIleLeuileAsnAlaArgIleArgGluGluThrGluAlaGlyAlaGlyAla 450
QY 539 AAAAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATCTTAACCTTG 598
Db 451 MetLysAlaLeuAspIleGlyPheAsnLysAlaThrIleIleAspSerAsnVal 470
QY 599 ACTACAGTATTGGCCTCAGCACCTTCTTCTCTAGATACAGGCGCTATTAAAGGGTTT 658
Db 471 ThrThrLeuAlaGlyThrValLeuLeuPheThrPheGlySerGlyProValLysGlyPhe 490
QY 659 GCTTGACATTGATTTAGGAATTTTCTCTCAATGTTTACGGCTCTTTTCATGACTATAA 718
Db 491 AlaValThrMetMetLeuGlyIleGlyIleSerMetPheThrSerValThrValValArg 510
QY 719 TTTTCTTCATGCTGTGGATGAATAGACCCCAACATACACAGTTCATATGATGAATAAG 778
Db 511 LeuLeuMetArgGluValValValValArgArgLysMetLysLysLeuGluIleProSerLeu 530
QY 779 TTGCTGGGGATAAAGCAT-----GATTCTTGTAGAGGATGCAAAAAACTTTGG 826
Db 531 PheGlyLysValProGlnLeuProThrPheSerPheMetLysGly----- 545
QY 827 GCTGTTTCTGGAAGTGTGTTTCTTTTAGGTGCTGCTCGGTTTGGAGCTTGAAT 886
Db 546 -----ArgPheLeuAlaIleGlyPheSerAlaPheLeu 556
QY 887 TCCGTT-----TTGGGAATCGGATTTTAAA 910
Db 557 SerIleSerSerValIleLeuPhePheThrProGlyLeuAsnTyrGlyIleAspPheIle 576
QY 911 GGAGGGTATGCTTTTACCTTTAATCCAAAAGAGCATGGCATGCGCATGTTGCTCAATG 970
Db 577 GlyIleGlnValGluAlaValSerLysGlyLys-----IleAsnLeuProThrLeu 594
QY 971 CGTGGCAAACTTGTGCATAAACTACAGGAAGCTGCTTCTTCTTAGAGACTTCCGTTATT 1030
Db 595 Arg-----GlnSerLeuGluGluLeuAsnLeuGly-----GluValAlaLeu 608
QY 1031 CAAACATTTCGATCTTCAGAAAAAGATCAAAATCTATTTTAGTGATATAA----- 1078
Db 609 GlnAspPheGlyGlyGlyGlnSerValLeuIleArgValGlnArgGlnProGlyGlyGlu 628
QY 1079 -----GCTTTAAGCTATACAGCAGATACGAGCTCTCTCTCAAAAATTA 1123
Db 629 GlnAlaGlnThrValAlaLeuAsnLysValLysAspAlaValThrThrAlaIleProGly 648
QY 1124 ACGATCATGAGCTGGCGTATTGTCGGGATTGTTGTCCAGAAACAGGCGCTAGATTCTCTAC 1183
Db 649 AlaSerMet----- 651
QY 1184 GGAACCTCTAAACGAACGCAAAATTTTGTCTCAAGTAAGACAGCAAACTATCGAAGAAA 1243
Db 652 -----GluArgThrGluValValGlyProThrValSerGlyGluLeuAlaArgSer 668
QY 1244 ATGCGTTATCAGCGGACCATCGGGCTTTTAGGA-----GCTTTGGCAATCATCTTG 1294
Db 669 -----GlyPheLeuAlaValAlaLeuAlaMetLeuAlaIleLeu 681

```

QY 1295 CTCTATGGAATTCGGCTTTGAATGCAATATGCTTCAGTCCGATGCGCTTATTT 1354  
 |||||  
 DB 682 LeuYrllrletrpPheArxPheglultrpHisPheAlaValGlyAlaIleAlaValIleu 701  
 QY 1355 CATGACCTTTGGGACTGTCGACACTCTGTTTATAGACATTTCTTTGGAAGAAATT 1414  
 |||||  
 DB 702 LeuAspIlethrLysThr-----ValGlyPhePheAlaIleuThrGlyLe 716  
 QY 1415 CAATATGATTTGCAAGCCATGTCGCTTTAATGACTGATTTGGGGTATTCATTAACAT 1474  
 |||||  
 DB 717 AspPheAlaIleuThrAlaIleAlaIleAlaIleuThrMetIleGlyTyrSerValAsnAsp 736  
 QY 1475 ACTTTGATCATTTTGTATGCTATTTGTTGAAGATCCCAAGCAACTGTTTAACTTATG 1534  
 |||||  
 DB 737 LysValValValValTyrAspArgMetArgGluAsnLeuArgLysTyrLysSerMetProhe 756  
 QY 1535 CATGTTTATGATGATGAGCCCTTCAAAAGACGTTGACGCGCAAGGATATGCAACACT 1594  
 |||||  
 DB 757 SerAspLeuIleAspMetSerIleAsnGlnValIleAlaArgCysIlePheThrSerAla 776  
 QY 1595 ACAACTATATGACGTTTGTCTAATGCTTTGTTTATAGCGCGCTCTGCTCTTATATTT 1654  
 |||||  
 DB 777 AlaThrAlaLeuSerLeuValProMetAlaIleTgpglyGluAlaValAlaGlnSerPhe 796  
 QY 1655 GCATTTATATGACCATGAGGATTTCTTATGAACTTTATGCTCTTATATATGACACA 1714  
 |||||  
 DB 797 AlaThrProMetIlePheGlyValIleValAlaThrThrSerSerIleTyrIleGly 816  
 QY 1715 CCTCGTGTGTTGTTATGTCGCGT 1738  
 |||||  
 DB 817 ProIleLeuLeuPheLeuSerArg 824

## RESULT 10

032047 PRELIMINARY; PRT, 737 AA.

AC 032047, 01-JAN-1998 (TREMblrel. 05, Created)

DT 01-AUG-1999 (TREMblrel. 11, Last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

DE SECDF protein (Protein-export membrane protein).

GN SECDF OR SECDF.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OC Bacillaceae; Bacillus.

OX NCBI\_TaxID=1423;

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RC MEDLINE=98044033; PubMed=9694879;

RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,

RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,

RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

RA Ertian K.D., Errington J., Fabret C., Ferrari B., Foulger D.,

RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,

RA Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,

RA Halpern H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,

RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudga B., Pak S.H.,  
 RA Paro V., Pohl T.M., Portetelle D., Potworlik S., Prescott A.M.,  
 RA Prescian E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Solio B.,  
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vanlier F., Vassartoli A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
 RA "The complete genome sequence of the gram-positive bacterium Bacillus  
 RT subtilis.";  
 RL Nature 390:249-256(1997).  
 RN [3]  
 RP SEQUENCE OF 167-737 FROM N.A.  
 RC STRAIN=168;  
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.,  
 RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF024506; AAC31122.1; -.  
 DR EMBL; Z99118; CAB14724.1; -.  
 DR InterPro: IPR001036; Acfrivin res.  
 DR InterPro: IPR003335; SecD\_SecF.  
 DR Pfam; PF02355; SecD\_SecF; 1.  
 DR PRINTS; PR00702; ACRIPLAVINRP.  
 DR TIGRFAMs; TIGR00916; 2A0604801; 2.  
 DR TIGRFAMs; TIGR01129; secD; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 737 AA; 81653 MW; 54C397E939779177 CRC64;

## Alignment Scores:

Pred. No.: 5.5e-49 Length: 737  
 Score: 678.50 Matches: 169  
 Percent Similarity: 50.43% Conservative: 125  
 Best Local Similarity: 28.99% Mismatches: 192  
 Query Match: 20.49% Indels: 97  
 DB: 16 Gaps: 15

US-09-868-987-1 (1-1864) x 032047 (1-737)

QY 104 GTCCAGAGCCCTATTAAAGTCCATGAAA--AATCATGCCAGTGTCCAGGAAA 160  
 |||||  
 DB 197 ValSerAlaProAsnValSerGlnLeuAsnThrThrAspValLysIleGlnGlyHis 216  
 QY 161 TTATACCAAGCGTGAAGTGAGCAAACTGCGCTCAGATTAAATGTGAGGAGATGCTTTT 220  
 |||||  
 DB 217 PheThrAlaGlnGlnAlaLysAspLeuAlaSerIleLeuAsnAlaGlyAlaLeu----- 234  
 QY 221 GTTCCGAGGTTCTCAGTGAAGAG-----ACGATCTCTTGTGATCTTGGAAAAA 271  
 |||||  
 DB 235 ---ProValLysLeuThrGlnLysTyrSerThrSerValGlyAlaGlnPheGlyGln 253  
 QY 272 CAATGTACACAGGACATTAATCTCAGATGCTGTGGCTTGCAAGCTTATGTTTATG 331  
 |||||  
 DB 254 AlaLeuHisAspThrValPheAlaGlyIleValaGlyIleAlaIleIlePheLeuPheMet 273  
 QY 332 AGCTATATATATGATTTGAGAGCGCTATCGCTCGGAGCATGTTCTTGAATCTTTTG 391  
 |||||  
 DB 274 LeuPheTyrThrArgLeuProGlyLeuIle-----AlaValIleThrLeuSerVal 290  
 QY 392 CTATGTCGGAGCTGTACAG-----TATTGATGCGGACATCACCTTGTACGA 442  
 |||||  
 DB 291 TyrIleTyrIleThrLeuGlnIlePheAspTyrPheAsnAlaValLeuThrLeuProGly 310  
 QY 443 CTCCTGGAGATGTTCTTCTGATAGGAGATGCGCTATGATGCAATGTTCTTGTATGAA 502  
 |||||  
 DB 311 IleAlaAlaLeuIleLeuGlyValGlyMetAlaValAlaAspAlaAsnIleIleThrGln 330  
 QY 503 AGAATCGAGAGGAATTTTATGTTCTCAAGTGTAAAGTGTAAATGTGTAAGAAAAGATAT 562  
 |||||  
 DB 331 ArgIleLysGlnGlnLeuLysLeuGlyLysSerValArgSerAlaPheArgSerGlyAsn 350





```

Db 270 LeuysserglyalaleuProilleysileValThralaaleuargserilleglyPro 289
Qy 257 GATCTGGGAAAAAACAATGACACAGGCAATTCAGCAGATGCTGGCTGGCAATG 316
Db 290 SerleuglylaapalaleargserglyalaleuValleuValleuValleuValleu 309
Qy 317 CTTATTTTGGATGAGGCTATATAT---AGATTGGAGCGCTCATGCTTCGGAGCT 373
Db 310 ValheValMeleuPheleuPheleuPheleuPheleuPheleuPheleuPheleu 329
Qy 374 GTTCTTCGAAATCTTTCTTATCTGGAGCAGCTTCAGTATTTGGATGCCCACTCAC 433
Db 330 LeuuePheSerleillelleleuglylleuuglylleuuglylleuuglylleuugly 349
Qy 434 TTGTACGACTCGCTGGATGTTCTTCTGATGGAGTGGCGGATGATGCAATGTTCTT 493
Db 350 LeuProglylleuValleuValleuValleuValleuValleuValleuValleuValleu 369
Qy 494 GTATTGCAAAATCCGAGAGATTTTATTTCTCAAGCTTTAAATAATCTGTAGAA 553
Db 370 SerPheglyargileuglyleuValleuValleuValleuValleuValleuValleu 389
Qy 554 AAAGATATACCAAGGCTTTGGAGCATTGTTGATTCATGCTTACATGATGATGCGC 613
Db 390 AlaelylyrghuhssetThralaalleuapValasnaHiseHiseuSer 409
Qy 614 TCAGCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 673
Db 410 AlaaleuValleuValleuValleuValleuValleuValleuValleuValleuValleu 429
Qy 674 TTAGGAATTTCTCTTCAATGTTTACGCTCTTTCATGACATAATTTTCTTCTGCTG 733
Db 430 IleglyValleuValleuValleuValleuValleuValleuValleuValleuValleu 448
Qy 734 TGGATGATATAGACCAACATACACAGTTGATGATGATGATGATGATGATGATGATG 793
Db 449 TTPeulaaglnhargyProhsn-----WeseAlaProglntPheleuValleu 464
Qy 794 CAT-----GATTCTTGAGAGATGACAAAACCTTGGCGCTTCTTCTTCTTCTTCTT 844
Db 465 HisThrHisPheaspPheMetleuValleuValleuValleuValleuValleuValleu 484
Qy 845 TTTCTTTAGGTTGCTGCTCTCGGCTTGGAGCCTGGAATCCGTTTGGGAATGAT 904
Db 485 AlaaleuValleuValleuValleuValleuValleuValleuValleuValleuValleu 502
Qy 905 TTTAAAGAGGATNGCTTACCTTATCCAAAGAGCATGCGATGCGATGCTTCT 964
Db 503 PheAlaProglyThrThleuThralaargValaspaarglnValThrThleu----- 520
Qy 965 CAATGCGTGGCAAGTGTGATTAACATACAGAGCTGCTTCTTCTTCTTCTTCTTCTTCT 1012
Db 521 GlnleuValleuValleuValleuValleuValleuValleuValleuValleuValleu 536
Qy 1013 -----CTTACGACTTCCGATTCACCAATTTGGATCTTCTTCTTCTTCTTCTTCT 1057
Db 537 GlnSerAlaThrIleGlnHargspThrThrProglynglnGlnHargspThrVal 556
Qy 1058 AAAATCTTTTATGATTAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTT 1117
Db 557 LysValProglyleuValleuValleuValleuValleuValleuValleuValleuValleu 574
Qy 1118 AAATTAACGATCATGAGCTGGCGTATTTGTGGATTTGTTCAGAAACAGGCTTGAATTT 1177
Db 575 LysleuPro----- 577
Qy 1178 CTCTACGAAACTTAAACGAAACGAAATTTGGTCAAGCTTACGACGAACTTATCG 1237
Db 578 ---GlnGlynglnValleuValleuValleuValleuValleuValleuValleuValleu 596
Qy 1238 AAGAAATACGCTTATCAGCGACCATGCGGCTTTAGAGGCTTGGCAATCATCTTGCCTC 1297

```

```

Db 597 GlnlyThrIleTyralaValleuLeuGly-----LeuGlyLeuIleLeuVal 612
Qy 1298 TATGTGATTTGGCGCTTGAATGGCAATATGCTTTCAGTCCGATAGCGCTTAAATCAT 1357
Db 613 TyralaGlyPheArgPheaspPheIleMeGlyLeuGlySerIleIleAlaIleHis 632
Qy 1358 GACCTTTGGCTACCTGTGACAGTCTTGTATGACATTTCTTTTGAAGAAATTCAA 1417
Db 633 AspaVal-----AlaIleAlaMeGlyLeuPheSerleuLeuGlyLeuGly 647
Qy 1418 ATAGATTTGCAAGCATGCTGCTTATGATGATGATGATGATGATGATGATGATGATG 1477
Db 648 PheThrValaIleSerValaIleAlaIleLeuLeuThrIleGlyTySerleuAsnaPse 667
Qy 1478 TTGATCATTTTGTATGCTATTTCTGGAAGATCGCCAGCAACCTGTTACCCCTATGAT 1537
Db 668 IleIleValSerPaspArgIleArgGlnAsMeIleThrMetArgGlyHisSerTyArg 687
Qy 1538 GTTTTATGATATGATGCTTCAAAAGACGTTGACCGGACGATGATGACACAGCTACA 1597
Db 688 GlnIleValasnaIleAlaIleAsnGlnThrLeuSerArgThrValMetThrSerValSer 707
Qy 1598 ACTCTATCAGTTTGTATGCTTGTGTTATGAGCGCTCTCTCTTCTTATTTTGA 1657
Db 708 ThrMetLeuProleuIleSerleuLeuIlePheGlyGlyProValleuArgAsPheSer 727
Qy 1658 TTTTATGACATAGGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1717
Db 728 LeuIleLeuValleuValleuValleuValleuValleuValleuValleuValleuValleu 747
Qy 1718 CTGTTGTTGTTATGATGCTC-----CGTAAAGAAATGCTCA 1753
Db 748 LeuValValleuValleuValleuValleuValleuValleuValleuValleuValleu 761

RESULT 12
ID 08YGT1 PRELIMINARY; PRT; 775 AA.
AC 08YGT1;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Protein translocase subunit SEC2 / protein translocase subunit
DE SEC2.
GN BMEI1076.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1,
RX MEDLINE=20020109; Pubmed=1175668;
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lydids A., Reznik G.,
RA Jeklovsk I., Larsen N., D'Souza M., Berni A., Mazur M., Goldsman E.,
RA Seltov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haevelorn R., Kyrides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).
DR EMBL: AE009547; AAL52257.1; -.
DR InterPro: IPR000408; Reg. chr. condens.
DR InterPro: IPR003335; SecD_SecF.
DR Pfam: PF02355; SecD_SecF; 1.
DR TIGRFAMs: TIGR00916; 2A0604s01; 2.
DR TIGRFAMs: TIGR00966; 3a0501s07; 1.
DR TIGRFAMs: TIGR01129; secD; 1.
DR PROSITE: PS00626; RCCL_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 775 AA; 83880 MW; 5BF72B71588A25FE CRC64;

Alignment Scores: 7.25e-47 Length: 775
Pred. No.:

```



Score: 653.50 Matches: 166  
Percent Similarity: 52.49% Conservativity: 140  
Best Local Similarity: 28.47% Mismatches: 204  
Query Match: 19.73% Indels: 73  
DB: 16 Gaps: 15

US-09-868-987-1 (1-1864) x Q8VGT1 (1-775)

```
QY 38 ACTGCTAATGACAAATATCTGCAACCGTGGATGGCTATGGCTGTAGTGAATGACGGT 97
    |||||
Db 219 ThrAlaGlnGlyAsnGluAsnSer-----PheAlaIleValValAspAsn 233
QY 98 TATATGCTCAGCACCCTATTTTAAACGTCCTCCATTG---AAAAATCATGCCAGTGTCTCA 154
    ::|||
Db 234 GlnValSerAlaProThrValSerGlyProLeuAspThrSerGluLeuGlnIleGlu 253
QY 155 GGGAAATTTACCCACCGTGAAGTCAGCAAACTCGCCCTCAGATTTTAAATCTGGAGCGCATG 214
    |||||
Db 254 GlyAlaPheAspLeuGlnAlaAsnMetAlaValValLeuArgSerGlyAlaLeu 273
QY 215 TCTTTTGTCCCGAGGTTCTCAGTGAAGACGATCTCTCTGATCTTGGGAAAAACAA 274
    |||||
Db 274 ProGlnAlaValThrValLeuGluArgThrIleAlaSerAlaLeuGlyGluAspTyr 293
QY 275 TGTACACAGCATATCTCAGCATGCTGTGGCTTGGCAATGCTATTGTTTGTATGATGACG 334
    ::|||
Db 294 AlaSerAlaAlaValLeuAlaAlaLeuAlaAlaValValGlyLeuPheMetVal 313
QY 335 GTATATTATAGATTGAGGGCGTCATCGCTTCGGAGCTGTCTCTCAATCTTTTCTT 394
    |||||
Db 314 LeuSerTyrGlyIleLeuGlyValIleAlaLeuValAlaLeuValAlaValIleIle 333
QY 395 ATCTGGCGAGCTCTACAGTATTTGGATCGCGCACTCACCTTGTGAGCACTCGCTGGGATT 454
    ::|||
Db 334 LeuThrAlaValLeuSerLeuIleGlyAlaSerIleSerLeuAlaSerIleAlaGlyLeu 353
QY 455 GTTCTTGTATGAGGATGGCGTGAAGTCAAAATGTTCTTGTATTCGAAAGATCCGAGAG 514
    |||||
Db 354 ValLeuThrIleGlyLeuAlaValAlaAspAlaHisIleLeuIleTyrGluArgValArgGlu 373
QY 515 GAATTTTATGTCTCAAGCTTTAAATACTCTAGAAAAGGATATACCAAGGCTTTT 574
    ::|||
Db 374 AspArgArgIysGlyTyrSerValValGlnAlaMetGluSerGlyPheTyrArgAlaLeu 393
QY 575 GGAGCCATTTTGTATCTAATCTGACTACAGTATTTGGCTCAGCACTCTTTTCTCTCTA 634
    |||||
Db 394 SerThrIleValAspAlaAsnLeuThrThrLeuIleAlaAlaLeuValLeuPheLeuLeu 413
QY 635 GATACAGGCTATTAAGGTTTGTGTTGACATGATTTTGAATTTTCTTCTTCAATG 694
    ::|||
Db 414 GlySerGlyThrValHisGlyPheAlaLeuThrValAlaIleGlyIleGlyThrLeu 433
QY 695 TTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAAT-----AAGACC 748
    |||||
Db 434 PheThrThrLeuThrPheThrArgLeuLeuIleAlaGlnTrpValArgThrAlaLysPro 453
QY 749 CAACATACACAGTTGCATATGATCAATAAGTTCTGGGATAAAGCAT---GATTTCTTG 805
    ::|||
Db 454 LysGluValProLysArgArgLeuLysLeuValProThrValThrHisIleProPheMet 473
QY 806 AGAGGATGCAAAAACCTTGGCGCTTCTGGAAGTGTCTTTCTTTAGGTGTC----- 859
    |||||
Db 474 Arg-----LeuGlnPheValThrLeuGlyIleSerValLeuAlaCysAlaIle 489
QY 860 -----GTTGCTCTCGGTTTGGAGCCCTGGAATTCGTTTGGGAATGAT 904
    |||||
Db 490 ValValAlaLeuPheValAsnIleGlyPheAsn-----TyrGlyIleAsp 504
QY 905 TTTAAGAGGGGTATGCTTTTACCTTTTAAATCCAAAA-----GAGCATGGCATCAGCGAT 958
    ::|||
Db 505 PheArgGlySerMetValGluLeuGlnAlaArgAsnGlyAspAlaAsnLeuGluAsp 524
QY 959 GTTGCTCAAAATGCGTGCAAAAGTTGTGCATAAACTACAGGAAGTGTCTTTCTCTAGA 1018
```

```
Db 525 IleAsnGlu-----ArgLeuAlaGluLeuAsnIleAspSer--- 536
QY 1019 GACTTCGCTATTCAACATTTGGATCTTTCAGAAAAGATCAAAATCTATTATTAGTGATAAA 1078
Db 537 -----Ala 537
QY 1079 GCTTTAAGCTATATAAGCAGATACGAGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1138
    ::|||
Db 538 ArgValLeuProAlaLysSerProArgSerAlaLeuVal-----IleIleGlySer 554
QY 1139 CGTTATTGTGG-----ATTGTTGTCTAGAACACAGCCTAGATTCTC 1180
    ::|||
Db 555 GlnGluValGlyAspAspAlaGluGlnThrValAlaValLysLeuArgGlyGluPheGlu 574
QY 1181 TACGAAACTCT---AAACGAAACGCAAAATTTTGGTCAAAAGTAAACGACGAACTACTCG 1237
    ::|||
Db 575 GlnAspTyrSerPheGlnArgValAspValValGlyProThrValSerGluGlnLeuSer 594
QY 1238 AAGAAATGCGTTATACAGCGCACCATCGCGCTTTTAGAGCTTTTGGCAATCATCTTCTC 1297
    ::|||
Db 595 ArgAlaGlyValLeuAlaValIleLeuSerLeuIleGly-----IlePheIle 610
QY 1298 TATGTGAGTTTGGCTTTGATGGAATATGCTTTCAGTCCGCTATCGCTTTAATTCAT 1357
    |||||
Db 611 TyrValTrpPheArgPheArgTrpGlnLeuAlaLeuGlyAlaValLeuSerThrLeuHis 630
QY 1358 GACCTTTTGGCTACCTGTGCAGCTCTGTTTATAGCACATCTCTTTTGAAGAAATTCAA 1417
    ::|||
Db 631 AspVal-----ValIleLeuSerGlyMetPheIleValPheArgMetGlu 645
QY 1418 ATAGATTGCAAGCCATTTGCTTTAATGACTGTATTTGGGTATTCATTAAACAATACT 1477
    ::|||
Db 646 PheAsnLeuTrpSerValAlaAlaIleLeuThrIleIleGlyTyrSerLeuAsnAspThr 665
QY 1478 TTGATCATTTTGTATGCTATTCGTGAAGATCGCCACGAACTGTTTACCCCTATGCAT 1537
    ::|||
Db 666 ValValIleTyrAspArgValArgGluAsnLeuArgTyrLysSerAlaProLeuPro 685
QY 1538 GTTTTGTAGTAAATGATGCCCTTCAAAAGACGTTTCAGCGCGCTCTCTCTCTCTCTCT 1597
    ::|||
Db 686 AlaIleLeuAspAlaSerIleAsnGlnThrLeuSerArgThrLeuLeuThrSerPheVal 705
QY 1598 ACTCTATCAGTTTGTATGCTTTTGTATAGCGGCTCTCTCTCTCTCTCTCTCTCTCTCT 1657
    ::|||
Db 706 ThrPheLeuAlaHisValProLeuTyrAlaPheGlyGlySerGluIleArgMetPheAla 725
QY 1658 TTATTATGACCATAGGATCTCTTAGGAACTTTATCGTCTCTTTATATTATGACCACT 1717
    ::|||
Db 726 LeuAlaLeuSerValGlyIleIleValAlaSerTyrSerIlePheIleAlaAlaPro 745
QY 1718 CTGTTGTTG 1726
Db 746 LeuLeuVal 748
```

RESULT 13

```
Q9ZBS8 PRELIMINARY; PRT; 795 AA.
AC Q9ZBS8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative SecDF protein-export membrane protein.
GN SC06160 OR SC1A9.24C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
```



```
QY 1403 TTGAAGAAATTCATATAGATTGCAAGCCATTGGTCTCTTAATGACTGTATTGGCGTAT 1462
Db 645 TrpLeuGlyLysProIleAspGlyValPheLeuAlaLeuLeuThrValIleGlyTyr 664
QY 1463 TCATTAAACCAATACCTTGATCATTTTGGATCGTATTCTGTA-----GATCGC 1510
Db 665 SerValAsnAspSerValValLeuPheAspArgIleArgGluLeuLeuGlyLysGluArg 684
QY 1511 CAAGCGAACCTGTTTACCCTATGATGATGTTTATAGTTAATGATGCCCTTCAAAAGACGTTT 1570
Db 685 LysAla-----ProPheAspArgLeuThrAsnAspAlaIleLeuGlnThrLeu 700
QY 1571 AGCGCACGCTATGACAAACAGCTACAACTCTATCAGTTTGTAAAGCTTTGTTTATA 1630
Db 701 ProArgThrValAsnThrGlyMetGlyAlaValLeuIleLeuAlaSerLeuAlaIleLeu 720
QY 1631 GCGCGCTCTGCTGCTTAAATTTGCACTTATATGACCATAGGATCTCTTAGGAAC 1690
Db 721 AlaAspSerLeuThrAspPheAlaLeuAlaLeuIleGlyValGlyValGlyThr 740
QY 1691 TTATCGCTCTTTATATGACACACCTCTGTTGTTGTTTATGTCCTGCGTAAGAAATCGC 1750
Db 741 TyrSerSerValPheThrAlaSerProLeuAlaIleGluLeuHisAsnArgAspThrGly 760
QY 1751 TCAAAA 1756
Db 761 SerArg 762

RESULT 14
Q92B17 PRELIMINARY; PRT; 754 AA.
AC Q92B17;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein lin1562.
GN lin1562.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Ruiniok C., Amend A.,
RA Baquero P., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Feihl H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkut G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
EMBL: AL596169; CAC96793.1; -.
DR L181818; LIN01562; -.
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF; 1.
DR TIGRFAMs; TIGR00916; 2A0604s01; 2.
DR TIGRFAMs; TIGR01129; secD; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 754 AA; 82546 MW; 047630364532DD4E CRC64;
```

## Alignment Scores:

Pred. No.:	4,52e-44	Length:	754
Score:	620.50	Matches:	162
Percent Similarity:	47.13%	Conservative:	117
Best Local Similarity:	27.36%	Mismatches:	222
Query Match:	18.73%	Indels:	91

```
DB: 16 Gaps: 13
US-09-868-987-1 (1-1864) x Q92B17 (1-754)
QY 98 TATATGGTCAGCAGCGCTATTTTAAACGTCCTCCATTTGAAATCATGCCAGTGCTCTCAGGG 157
Db 196 TyrLeuSerAlaProAsnValSerSerValLeuAspThrAspLysValGluIleSerGly 215
QY 158 AAATTATCCACCGCTGAAGTCAGCAAACTCGCTCAGATTTAAATCTCGAGCGATGCT 217
Db 216 SerPheThrThrGluGluAlaLysAspLeuAlaGluLeuLeuAsnSerGlyAlaLeuPro 235
QY 218 TTTGTTCCCGAGGTTCTCAGTCAGAGACGATCTCTCTGATCTTGGGAAAAACAATGT 277
Db 236 ValLysMetThrGluValTyrSerThrSerValGlyAlaGlnPheGlyGlnAspAlaLeu 255
QY 278 ACACAGGCAATATCTCAGCATGCTGTGGCTTCGCAATGCTTATTGTTTGGATGAGCGTA 337
Db 256 GlnGluThrIleLeuAlaGlyIleIleGlyValMetAlaIlePheIlePheMetMetAla 275
QY 338 TATTATAGATTTTCGAGCGCTCATCGCTTCGGAGCTGTTCTCTGAAATCTTTTCTTATC 397
Db 276 ValTyrArgLeuProGlyValIleAlaSerIleThrLeuValAlaTyrThrTyrLeuVal 295
QY 398 TGGGACGCTCTACAGTATTTGGATGCGCCACTCACCCTGTGACGACTCGCTGGGATGTT 457
Db 296 LeuLeuIleLeuSerLeuLeuAsnAlaThrLeuThrLeuProGlyIleAlaGlyLeuIle 315
QY 458 CTGTGCTATGGGATGCGCTAGATGCAATGTTCTGTATTCGAAAGAATCCGAGAGGAA 517
Db 316 LeuGlyIleGlyMetAlaValAlaAspAlaAsnValIleThrTyrGluArgIleLysGluGlu 335
QY 518 TTTTATTTGCTCAAAAGTCCTTAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGG 577
Db 336 IleLysValGlyArgSerThrLysAlaAlaPheGluValGlyLysGluAlaPheArg 355
QY 578 GCCATTTTTCATTTCACTTACAGTATGCGCTCAGCTCAGCTCTCTTTCTTCTCTAGAT 637
Db 356 AlaIleLeuAspGlyAsnLeuThrThrLeuIleValAlaAlaValLeuPheTyrPheGly 375
QY 638 ACAGGCGCTATTAAAGGTTGCTTTGACATTTAGCAATTTTCTCTCAATGTTT 697
Db 376 ThrSerSerIleLysGlyPheAlaThrValLeuIleIleSerIleLeuValSerPheLeu 395
QY 698 ACGGCTCTTTTCATGATAAATTTTCTTCATGCTG-----TGATGAAT 742
Db 396 ThrAlaValTrpGlySerArgPheLeuLeuGlyLeuValLysSerAsnTrpLeuAsn 415
QY 743 AAG-----ACCAACATACACAGTTGCGATGATGATGATGATGATGATGATGATGAT 784
Db 416 AsnLysProGlyPhePheAlaValLysArgLysAspIleHisAsnLeuHisGluGlyIle 435
QY 785 GGATA-----AAGCATGATTTCTTGAGAGGATGAGAGGATGAGAGGATGAGAGGAT 820
Db 436 AsnSerPheSerLeuLysThrHisPheAspArgTyrAspPheValLysHisArgLeu 455
QY 821 CTTTGGGCTGTTCTGGAAGCTTTTCTTTAGTGTGCGTGTCTCTCGGCTTTTGGAGCC 880
Db 456 PheLeuSerIlePheAlaAlaIleValIleValIleValIleValIleValIlePheLys 475
QY 881 TGAATTCCTTTTGGGAATGGAATTTTAAAGGAGGATGATGATGATGATGATGATGATGAT 940
Db 476 LeuAsn-----LeuGlyIleAspPheAlaSerGly-----ThrArg 487
QY 941 GAGCATGGCATCAGCATGTTGCTCAATCGCTGGCAAAAGTGTGTCATAAATACATACAGAA 1000
Db 488 AlaGluValThrAlaAsnGlnThrLeuThrGluThrGlnIleLysLysAspLeuAspAla 507
QY 1001 GCTGGTCTTTCTTAGAGACTTCCGGTATTTCAAAACATTTTGGATCTTCA----- 1048
Db 508 IleAspMetProSerAsnAspIleValPheGlnGlySerGlySerLysThrAlaValVal 527
QY 1049 -----GAAAAAGATCAAAATCTATTTTAGTCAT 1075
```

```

Db 528 SerTyrIysGlyThrLeuSerClnAsnAspValAlaIlySphLeuAspTyrPheGluAsp 547
Qy 1076 AAAGCTTAACTATACAGAGATACAGAGCTCTCTCTAAATTAACATGATGAGC 1135
Db 548 Lys-----TyrLysHISgluProSerIleSerThrValSerProThrVal----- 562
Qy 1136 TGGCGTTATTGGGATGTGTGTCTGAGAAACAGCGCTAAGATTCTCTAGGAACTCTAAA 1195
Db 563 -----GlyLysGluLeu 566
Qy 1196 CGAAACGAAATTTGGTCAAGGTAAAGACAAACTATGAGAAATGCGTATACG 1255
Db 567 AlAllysAsnGlyPheTrpAla----- 573
Qy 1256 GCGACCATCGGGCTTTAGCA---GCTTTGGCAATCATCTGCTCTATGAGTTGGCGC 1312
Db 574 -----LeuGlyValAlaSerValLeuIleValLeuTyrIleAlaValArg 588
Qy 1313 TTGGAATGGCAATATGCTTTCAGTCCGTATGCGCTTAAATTCATGACCTTTGGCTACC 1372
Db 589 PheGluPheTyrMetGlyIleAlaIleAlaLeuSerLeuLeuPheAsp----- 604
Qy 1373 TGTGAGCTTGTATTATAGCACATTTCTTTGAAGAAATTCAAATGATTTGGCAAGCC 1432
Db 605 ---AlaPheIleIlePheIlePhePheSerValThrArgLeuGluValAlaSpheuThrPhe 623
Qy 1433 ATTGGTCTTAAATGATGCTATGTTGGGGTATTCATTAAACAATCTTATGATCTTTGAT 1492
Db 624 IleAlaIleAlaValLeuThrValIleGlyTyrSerIleAsnAspThrIleValThrAlaAsp 643
Qy 1493 CGTATTCGTGAA---GATCCGACAGCGCAACCTGTTACCCCTATGCAATGTTTAA----- 1543
Db 644 ArgIleArgAspIleSerMetLysMetGlnArgPheLysTrpLysGluGluIleAlaAsp 663
Qy 1544 ---GTTATGATGACCTTCGCAAGAGCTTCAGCCGACGATATGACAAAGCTCAACT 1600
Db 664 AlaValAlaValAlaLeuValGlnThrPheThrArgSerIleAsnThrIleLeuThrVal 683
Qy 1601 CTATCAGTTTGTATGATGCTTGTGTTTATAGCGGCGCTCTGCTGCTTAAATTTGCAATT 1660
Db 684 IlePheThrValLeuAlaValValLeuPheGlySerGluSerIleLeuAsnPheSerIle 703
Qy 1661 ATTATGACCATAGGAGATCTTCTAGAACTTATGCTCTTATATATGACACCACTCTG 1720
Db 704 AlaLeuLeuValGlyLeuValSerSerValPheSerSerIlePheMetAlaMetGlnLeu 723
Qy 1721 TTGTGTTTATGATGCTGCTGTAAGAAATGCTCAAAA 1756
Db 724 TrpTyrValPheLysAlaArgGlnLeuArgLysLys 735

RESULT 15
ID 08Y703 PRELIMINARY; PRT; 754 AA.
AC 08Y703;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein Imo1527.
GN Imo1527.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=BGD-E / SEROVAR 1/2a;
RX MEDLINE=21537279; PubMed=11679669; C., Rusniok C., Amend A.,
RA Glaser P., Frangeul L., Buchrieser T., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Baquero F., Cherouani F., Couve E., de Daruvar A., Deboux P.,
RA Charbit A., Cherouani F., Couve E., de Daruvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussart O.,
RA Entian K.-D., Fathi H., Garcia-del Portillo F., Garrido P.,

```

```

RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerster U., Kreft J., Kuhn M., Kunst F., Kurapack G.,
RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjati H.,
RA Nordstok G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
RT "Comparative genomics of Listeria species.";
RL Science 294:845-852(2001).
DR EMBL; AL591979; CAC9605.1; -.
DR ListIstic; LM001527; -.
DR InterPro; IPR003335; SecD_SecE.
DR Pfam; PF02355; SecD_SecE.1.
DR TIGRFAMs; TIGR00916; 2A0604s01; 2.
DR TIGRFAMs; TIGR01129; secD; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 754 AA; 82614 MW; F2BA63B501EB58B CRC64;

Alignment Scores:
Pred. No.: 6,68e-44 Length: 754
Score: 618.50 Matches: 166
Percent Similarity: 46.79% Conservative: 111
Best Local Similarity: 28.04% Mismatches: 224
Query Match: 18.67% Indels: 91
DB: Gaps: 13

US-09-868-987-1 (1-1864) x 08Y703 (1-754)
Qy 98 TATATGCTCAGACAGCCCTATTTAAACGTCCTCAATGAAATCATGCCAGTCTCAGG 157
Db 196 TyrLeuSerAlaProAsnValSerSerValLeuAspThrAspLysValGluIleSerGly 215
Qy 158 AAATTTCACCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 217
Db 216 SerPheThrArgGluAlaValAspLeuAlaValLeuLeuAsnSerGlyAlaLeuPro 235
Qy 218 TTTGTTCCGAGGTTCTCAGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 277
Db 236 ValLysMetLysGluValTyrSerThrSerValGlyAlaGlnPheGlyGlnAspAlaLeu 255
Qy 278 ACACAGGACATATCTCAGACAGTGTGCTTGGCAATGCTTATGTTGATGAGCGTA 337
Db 256 GlnGluThrIleLeuAlaGlyIleIleGlyValIleAlaIlePheIlePheMetMetAla 275
Qy 338 TATATGATTTGAGAGCGCTATGCTTGGAGAGCTTCTCTGATCTTGGATCTTTGCTATC 397
Db 276 ValTyrArgLeuProGlyValIleAlaSerIleThrLeuValAlaTyrThrTyrLeuVal 295
Qy 398 TGGGACCTCTACAGATTTGGATGCGCACTCTGTCAGGACCTGCGTGGATGTT 457
Db 296 LeuLeuIleLeuSerLeuLeuAsnAlaThrLeuThrLeuProGlyIleAlaGlyLeuIle 315
Qy 458 CTTCATGAGGAGATGCGCGATGATGCAATGCTTGTATTCGAAAGATCCGAGAGAA 517
Db 316 LeuGlyIleGlyMetAlaValAlaAspAlaValIleThrTyrGluArgIleLysGluGlu 335
Qy 518 TTTTATGCTCAAGTCTTAAAAATCTGTAGAAAAGATPACCAAGCTTTTGA 577
Db 336 IleLysValGlyArgSerThrLysAlaIlePheIleValGlyLysGluAlaPheArg 355
Qy 578 GCAATTTTGTATCTACCTGATGATGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGAT 637
Db 356 AlaIleLeuAspGlyAsnLeuThrThrLeuIleValAlaIleValAlaLeuPheTyrPheGly 375
Qy 638 ACAGGCTTATTAAGGATGCTTTCATGATGATGATGATGATGATGATGATGATGATGAT 697
Db 376 ThrSerSerIleLysGlyPheAlaThrValLeuIleIleSerIleLeuValSerPheLeu 395
Qy 698 ACGGCTCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 742
Db 396 ThrIleValTrpGlySerArgPheLeuGlyLeuLeuValLysSerAsnTrpLeuAsn 415
Qy 743 -----AAGCAACCAATATACAGCTTGCATATGATG 772

```

```

Db 416 AsnLysProGlyPhePheAlaValLysArgLysAspIleHisAsnLeuHisGluGlyLeu 435
QY 773 AATAAGTTCTGGGGATAAAGCAT-----GATTTCTTGAGAGGATGCACAAAAA 820
Db 436 AsnSerPheSerLeuLysThrHisPheAspArgPheAspPheValLysHisArgLeu 455
QY 821 CTTTGGGCTGTTCTGGNAGTGTTTCTTTTCTTTAGTTCGCTGCTCTCGGTTTGGAGCC 880
Db 456 PheLeuSerIlePheAlaAlaIleValLysGlyLeuValIleLeuSerIlePheArg 475
QY 881 TGAATATCCGTTTGGGAATGGATTTAAAGAGGGGTATGCTTTTACCTTTAAATCCAAAA 940
Db 476 LeuAsn-----LeuGlyIleAspPheAlaSerGly-----ThrArg 487
QY 941 GAGCATGGCATCAGCGATGTTGCTCAATCGTGGCAAAAGTTGTGCATAACTACAGAA 1000
Db 488 AlaGluValThrAlaAsnGlnThrLeuThrGluThrGlnIleLysAspLeuAspThr 507
QY 1001 GCTGGTCTTTCTTAGAGACTTCGGTATTCAACATTTGGATCTTCA-----1048
Db 508 IleAspMetProSerAspAspIleValPheGlnGlySerGlySerLysThrAlaValVal 527
QY 1049 -----GAAAGATCAAAATCTATTTTAGTGAT 1075
Db 528 SerTyrLysGlyThrLeuSerGlnAsnAspValAlaLysPheLysAsnTyrPheGluAsp 547
QY 1076 AAAGCTTTAAGCTATACATAGCAGATAGAGCCCTCTCTCTTAAATTAACGATCATGAGC 1135
Db 548 Lys-----TyrLysHisGluProSerIleSerThrValSerProThrVal-----562
QY 1136 TGGCGTTATTGCGGATGTTGTGAGAAACAGGCTAGATTCTCTACGGAAATCTTAAA 1195
Db 563 -----GlyLysGluLeu 566
QY 1196 CGAAACGCAAAATTTTGTCAAGGTAAGCAGCAAACTATCGAAGAAATGCGTTATCAG 1255
Db 567 AlaLysAsnGlyPheTrpAla-----573
QY 1256 GCGACCATCGGGCTTTTAGGA---GCTTTGCAATCATCTTGTCTATGTGAGTTTGGCG 1312
Db 574 -----LeuGlyValAlaSerValLeuIleValLeuTyrIleAlaValArg 588
QY 1313 TTTCGAATGGCAATATGCTTTTCAGTCCGATGCGCTTAAATCATGACCTTTTGGCTACC 1372
Db 589 PheGluPheTyrMetGlyIleAlaAlaIleLeuSerLeuLeuPheAsp-----604
QY 1373 TGTCCAGTCTTGTTTATAGCACATTTCTTTTGAAGAAATTCAAATAGATTTCGAAGCC 1432
Db 605 ---AlaPheIleIlePhePheSerValThrArgLeuGluValAspLeuThrPhe 623
QY 1433 ATTGGTCTTTTAAATGACTGATTTGGGGTATTTCATTAAACAATCTTTGATCATTTTGTAT 1492
Db 624 IleAlaValLeuThrValIleGlyTyrSerIleAsnAspThrIleValThrAlaAsp 643
QY 1493 CGTATTCTGTGAA---GATCGCAACGCACTGTTTACCCTCATGCTGTTTAA-----1543
Db 644 ArgIleArgAspIleSerMetLysMetGlnArgPheLysThrLysGluGluIleAlaAsp 663
QY 1544 ---GTTAATGATGCCCTTCAAGAGCGTTTCAGCCGAGTAATGACACAGCTACTCAACT 1600
Db 664 AlaValAsnLysAlaLeuArgGlnThrPheThrArgSerIleAsnThrIleLeuThrVal 683
QY 1601 CTATCAGTTTGTAAATGCTTTTGTGTTTATAGCGGCTCCTCTGCTTTTAAATTTTGCATTT 1660
Db 684 IlePheThrValLeuAlaLeuValLeuPheGlySerGluSerIleLeuAsnPheSerIle 703
QY 1661 ATTATGACCATAGGATCTTCTAGGAACTTTATGCTCTCTTTATATATGACCACTCTG 1720
Db 704 AlaLeuLeuValGlyLeuValSerSerValPheSerSerIlePheMetAlaMetGlnLeu 723
QY 1721 TTGTTGTTTATGTCCTGTAAGAAATCGCTCAAAA 1756
Db 724 TrpTyrValPheLysAlaArgGlnLeuArgLysLys 735

```

## RESULT 16

```

Q98PRO PRELIMINARY; PRT; 873 AA.
ID AC Q98PRO;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-export membrane protein SECD.
GN MYPV_6590.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
NCBI_TaxID=2107;
RN [1]
RX MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Sanson D., Galisson F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis."
RL Nucleic Acids Res 29:2145-2153 (2001).
DR EMBL; AL45565; CAC13632.1; -.
DR Mypulist; MYPV_6590; -.
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF_1.
DR TIGRFAMs; TIGR00916; 2A0604801; 2.
KW Complete proteome.
SQ SEQUENCE 873 AA; 97332 MW; 02B06BB39DDF9231 CRC64;

```

## Alignment Scores:

```

Pred. No.: 9,23e-36 Length: 873
Score: 522.50 Matches: 143
Percent Similarity: 47.77% Conservative: 135
Best Local Similarity: 24.57% Mismatches: 195
Query Match: 15.78% Indels: 109
DB: 16 Gaps: 18

```

US-09-868-987-1 (1-1864) x Q98PRO (1-873)

```

QY 149 GTCTCAGGAAATTTACCCACCGTGAAGTAGAGCAAACTCGCTCAGATTTTAAATCTGGA 208
Db 298 IleThrGlyAsnLeuThrAlaSerSerAlaLysGlnLeuAlaLeuAspValAsnPheGly 317
QY 209 GCGATGCTTTTGTTCCTCGAGGTTCTCAGTGAAGAGACGATCTCTCTGATCTTTGGGAAA 268
Db 318 SerGlyAspTyrThrLeuGluLeuLeuSerSerArgPheValAspAlaThrLeuGlyThr 337
QY 269 AAACAATGTACACAAGGCATTTATCTCAGCATGCTGTGGCTTGGCAATGCTTATTGTTTG 328
Db 338 GlnSerPheAsnTyrAlaIleIleAlaGlyLeu-----IleSerPheIleValIleAla 355
QY 329 ATGAGCTATATTATAGATTGGA-----GGCGTCACTCGCTTGGGAGCTGTTCTTCTG 382
Db 356 LeuIleMetIleTyrAsnTyrGlyValLeuGlyIleAlaSerSerLeuSerIleGlyLeu 375
QY 383 AATCTTTTGTCTATCTGGGAGCTCTACAGTATTGTGATGGCCACTCACCTTGTCTAGGA 442
Db 376 TyrIlePheLeuSerLeuThrValPheThrTrpLeuAsnGlyGluTyrSerProAlaThr 395
QY 443 CTGCTCGGATTTGTTCTGCTATGGGATGGCGTAGATGCAATGTTCTTGTATTTCGAA 502
Db 396 IleAlaLeuValIleAlaIleIleGlyMetSerValAspAlaAsnIleIleThrPheGlu 415
QY 503 AGAATCCGAGGAGAAATTTTATTGCTCTCAAGTCTTAAAAATCTGTAGAAAAAGGATAT 562
Db 416 ArgIleLysGlnGlnLeuLysTyrGlyAspSerLeuLysAlaThrLysAsnAlaHis 435
QY 563 ACCAAGCTTTTGGAGCAATTTTGTATTCTTAAGTGTAGTACAGTATTGGCTCAGACTT 622
Db 436 ArgPheSerLeuSerThrIleLeuAspSerAsnIleThrThrLeuIleValSerGlyLeu 455

```



```

Qy 359 ATCGCTCGGAGCTGTTCTCTGAATCTTTTGGCTATCTGGGAGCTCTACAGTATTG 418
Db 266 ILeAlaAspLeuAlaIleIlePheGlyPheIleThrPheAlaCysLeuAsnPheIle 285
Qy 419 GATCGGCACCTCCTGCTCAGGAGCTGCTGGGATTGTTCTGCTATGGGAGTGGCGTA 478
Db 286 AspAlaThrLeuThrLeuProGlyIleAlaGlyPheIleLeuSerLeuGlyMetAlaVal 305
Qy 479 GATGCAATGTTCTGTATTCGAAAGATCCGAGAGCAATTTTATTGTCrCAAAAGTCTT 538
Db 306 AspAlaAsnValIleIlePheGluArgIleLysGluLeuArgPheGlyAsnSerIle 325
Qy 539 AAAAAATCTGAGAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATCTTAACCTG 598
Db 326 ArgAsnSerIleAspSerGlyPheAsnLysGlyPheIleAlaIlePheAspSerAsnLeu 345
Qy 599 ACTACAGTATGGCTCAGCAGCTCTCTTCTCTCCTAGATACAGGCGCTATTAAAGGGTTT 658
Db 346 ThrThrLeuIleIleThrIleLeuPheValPheGlyThrGlyProIleIleGlyPhe 365
Qy 659 GCTTTGACATGATTTAGGAAATTTCTCTCAATGTTTACCGCTCTTTTCATGACTAAA 718
Db 366 AlaValThrLeuAlaLeuGlyThrLeuAlaSerMetPheThrAlaIleThrValThrLys 385
Qy 719 TTTTCTTCATGCTGTGATGAAT 742
Db 386 ValLeuLeuLeuThrPheValAsn 393

RESULT 18
Q8ZH5 ID Q8ZH5 PRELIMINARY; PRT; 474 AA.
AC Q8ZH5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Protein-export membrane protein.
GN SEC OR ALL0121.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003581; BAB7645.1; -
DR InterPro; IPR000911; Ribosomal_L11.
DR InterPro; IPR003335; Secd_Secf.
DR Pfam; PF02355; Secd_Secf.1.
DR TIGRFAMs; TIGR00916; 2A0604s01.1.
DR TIGRFAMs; TIGR01129; secD.1.
DR PROSITE; PS00359; RIBOSOMAL_L11; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 474 AA; 50454 MW; 67ABEB4BE4A3D564 CRC64;

Alignment Scores:
Pred. No.: 1.86e-28 Length: 474
Score: 436.00 Matches: 89
Percent Similarity: 60.61% Conservative: 51
Best Local Similarity: 38.53% Mismatches: 85
Query Match: 13.16% Indels: 6
DB: 16 Gaps: 1

US-09-868-987-1 (1-1864) x Q8ZH5 (1-474)
Qy 59 GCAACCGTGGATGGCGTATGGCTGTAGTGTATGCGGTTATATGGTCAGCAGCCCTATT 118

```

```

Db 214 AlaGlyThrGlyArgSerIleGlyIlePheLeuAspAsnGluLeuIleSerAlaProAsn 233
Qy 119 TTAACCTGCCATTTGAAAAAT-----CATGCCAGTGTCTCAGGAAA 160
Db 234 ValGlyIleGluPheAlaSerThrGlyIleThrGlySerAlaValIleThrGlyArg 253
Qy 161 TTATCCACCCTGAGTGAAGCAAACTCCCTCAGATTAAATCTGAGGAGTCTCTTTT 220
Db 254 PheThrAlaGlnGlnAlaAsnAspLeuGlyValGlnLeuArgGlyGlyAlaLeuProVal 273
Qy 221 GTTCCCAGGTTCTCAGTGAAGAGACCATCTCTCTCATCTTGGGAAAAACAATGTACA 280
Db 274 ProValGluIleAlaGluIleArgThrValGlyAlaThrLeuGlyLysAspSerIleAsn 293
Qy 281 CAAGGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTATTGTTTGTGATGAGCGTATAT 340
Db 294 SerSerIleThrAlaGlyLeuGlyGlyLeuThrLeuValLeuIlePheMetValValTyr 313
Qy 341 TATAGATTTGGAGCGTCATCGCTTCGGAGCTGTTCTCTGAATCTTTTCTCTATCTGG 400
Db 314 TyrArgLeuProGlyLeuIleAlaAspIleSerLeuIleIleTyrAlaIleLeuThrTrp 333
Qy 401 GCAGCTCTACAGTATTTGGATGGCCACTCCTCCTGTCAGGACTCGCTGGGATGTTCTT 460
Db 334 AlaSerPheAlaLeuLeuGlyIleThrLeuThrLeuProGlyIleAlaGlyPheIleLeu 353
Qy 461 GCTATGGGATGGCGCTAGATGCAAAATGTTCTGTATTTCGAAAGAAATCCGAGAGGAATTT 520
Db 354 SerIleGlyMetAlaValAlaAspAlaAsnValLeuIlePheGluArgThrArgGluGluLeu 373
Qy 521 TTATGTCTCAAGCTTAAAAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCC 580
Db 374 GlnAlaGlyLysSerLeuTyrArgSerValGluSerGlyPheTyrArgAlaPheSerSer 393
Qy 581 ATTTTGTATCTAACTGACTACAGTATTTGGCTCAGCAGCTCTCTTTCTCTCTAGATACA 640
Db 394 IleLeuAspGlyAsnValThrValIleAlaCysAlaAlaLeuPheTrpLeuGlyAla 413
Qy 641 GGGCTATTAAAGGGTTTGTCTGACATTTGTTTACGAAATTTTCTCTCAATGTTTACG 700
Db 414 GlyLeuValLysGlyPheAlaLeuThrLeuAlaLeuGlyValAlaValSerMetPheSer 433
Qy 701 GCTCTTTTCATGACTAAATTTTCTCTCATGCTG 733
Db 434 AlaValThrCysSerArgThrLeuMetPheLeu 444

RESULT 19
Q9HX11 ID Q9HX11 PRELIMINARY; PRT; 620 AA.
AC Q9HX11;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Secretion protein SecD.
GN SEC OR PA3821.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).

```



DR EMBL: AE004799; AAC07208.1; -.  
 DR InterPro: IPR003335; SecD\_SecF.  
 DR Pfam: PF02355; SecD\_SecF; 1.  
 DR TIGRFAMs: TIGR00916; ZAO604s01; 1.  
 DR TIGRFAMs: TIGR01129; secD; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 620 AA; 67674 MW; BA8740DB46A099FD CRC64;

## Alignment Scores:

Pred. No.: 1,98e-27 Length: 620  
 Score: 424.00 Matches: 103  
 Percent Similarity: 53.71% Conservative: 49  
 Best Local Similarity: 36.40% Mismatches: 83  
 Query Match: 12.80% Indels: 48  
 DB: 16 Gaps: 5

US-09-868-987-1 (1-1864) x Q9FX11 (1-620)

```

QY 62 AACCGTGGATGGCGTATGCTGATG----- 88
DB 357 AsnValGlyArgSerMetAlaValAlaPheIleGluInLysProValThrArgTyrThr 376
QY 89 -----ATTGACGGT----- 97
DB 377 LysGlnMetValAspGlyValGluLysGluValAlaValProAlaPheLysGluGluLys 396
QY 98 TATATGCTACGACGCCCTATTTTAAAGCTCCCATTTGAAAAATCATGCCAGTCTCAGGG 157
DB 397 GlnIleIleSerLeuAlaThrIleGlnSerProLeuGlyAsnGlnPheArgIleThrGly 416
QY 158 AAATTTACCCACCGTGAAGAGAGCAACATCGCTCAGATTTAAATCTGAGCGCATGCT 217
DB 417 LeuAspGlyProGlyGlnSerSerGlnLeuAlaLeuLeuAlaGlyGlyLeuAla 436
QY 218 TTTGTTCCCGAGTTCCTCATGACAGACGATCTTCTGATCTTGGAAAAACAATGT 277
DB 437 AlaProMetCyrPheAlaGluGlnArgThrIleGlyProSerLeuGlyAlaAspAsnIle 456
QY 278 ACACAGGCAATTTATCTCAGCATGCTGTCGGCATTCCTATTTGTTGAGCGCTA 337
DB 457 AlaLysGlyIleAspAlaSerLeuThrPrgLysMetLeuPheValSerLeuPheIleIleVal 476
QY 338 TATATAGATTGGAGGGGTATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTATC 397
DB 477 IleTyrArgPhePheGlyValIleAlaThrValAlaLeuAlaPheAsnMetValMetLeu 496
QY 398 TGGCAGCTCTACAGTATTTGGATGCGCCACTCAGCTTGTGAGAGCTGCTGAGATTGT 457
DB 497 ValAlaLeuMetSerIleLeuGlyValaThrLeuThrLeuProGlyIleAlaGlyIleVal 516
QY 458 CTTCGCTATGGGATGCGCGTACATGCAAAATGTTCTTGTATTCGAAAAATCCGAGAGAA 517
DB 517 LeuThrMetGlyMetAlaValAspAlaAsnValLeuIlePheSerArgIleArgGluIle 536
QY 518 TTTTATATGTCGCAAGCTTAAAGCTGAGAAAAAGATATACCAAGGCTTTTGA 577
DB 537 LeuAlaAsnGlyMetSerValGlnArgAlaIleHisGluGlyPheAsnArgAlaPheThr 556
QY 578 GCCATTTTGGATTCTACTACAGTATTTGGCGCTCAGCACTTTTCTTCTCTAGAT 637
DB 557 AlaIleLeuAspAlaAsnLeuThrSerLeuLeuValGlyGlyIleLeuThrAlaMetGly 576
QY 638 ACAGGGCTATTAAGGGTTGCTTGAACATGATTTTGAAGATTTTCTCTTCAATGTTT 697
DB 577 ThrGlyProValLysGlyPheAlaValThrMetSerLeuGlyIleIleThrSerMetPhe 596
QY 698 ACGGCTCTTTCATGACATTAATTTTCTTCATGCTGCTGAGTAAGTAACCAACATATCA 757
DB 597 ThrAlaIleMetValThrArg----- 603
QY 758 CAGTTGATGATGATTAAGTTCGTGGGATTAAGCATGATTTTCTTGAGAGATGACAA 817
DB 604 -----AlaMetValAsnLeuIlePheGly---GlyArgAspPhe-----Lys 616

```

QY 818 AAATTTGG 826  
 DB 617 LysLeuThr 619

## RESULT 20

08XJ19 PRELIMINARY; PRT; 419 AA.

AC 08XJ19;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Protein-export membrane protein.  
 GN SBCD OR CPE1942.  
 OS Clostridium perfringens.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Clostridiales; Clostridiaceae; Clostridium.  
 NCBI\_TaxID=1502;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=13 / TYPE A.  
 RX PubMed=11792842;  
 RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,  
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,  
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
 RT flesh-eater";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001 (2002).  
 DR EMBL: AP003192; BAB81648.1; -.  
 DR InterPro: IPR003335; SecD\_SecF.  
 DR Pfam: PF02355; SecD\_SecF; 1.  
 DR TIGRFAMs: TIGR00916; ZAO604s01; 1.  
 DR TIGRFAMs: TIGR01129; secD; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 419 AA; 45126 MW; BAF29D24BF81652 CRC64;

## Alignment Scores:

Pred. No.: 2.1e-27 Length: 419  
 Score: 423.50 Matches: 87  
 Percent Similarity: 60.57% Conservative: 62  
 Best Local Similarity: 35.37% Mismatches: 96  
 Query Match: 12.79% Indels: 1  
 DB: 16 Gaps: 1

US-09-868-987-1 (1-1864) x 08XJ19 (1-419)

```

QY 65 CGTGATGGCGTATGCTGATGATGACGCTTATATGTCAGCACCCCTATTTAAAC 124
DB 173 LysGlyGlnLysIleAlaIleLysMetAspAsnGluThrLeuThrAspProValValAsn 192
QY 125 GTCCCATGAAAT---CATGCCAGTCTCAGGGGAAATTTACCCACCGTGAAGTACG 181
DB 193 AspIleIleSerAsnGlyGluAlaIleIleSerLysAsnArgSerMetGluGluAlaGlu 212
QY 182 AAATCCGCTCAGATTTAAATCTGAGCGATGCTTTGTTCCGAGTCTTCAGTGA 241
DB 213 LysValSerGlyIleIleAsnAlaGlyAlaLeuProValProValLysAlaValSerVal 232
QY 242 GAGACGATCTCTTGTGATCTTGGGAAAAACAATGTACACAGGCAATTTCTCAGCATGC 301
DB 233 GluThrValGlyAlaGlnLeuGlyAlaAsnAlaLeuProAsnAlaLeuLysAlaGlyAla 252
QY 302 TGTGCGTTCGCAATGCTTATTTGATGAGCGTATATATATGATTTGGAGCGCTCATC 361
DB 253 IleGlyValAlaIleIlePheLeuPheMetIleLeuTyrThrGValProGlyPheIle 272
QY 362 GCTTCGGAGCGCTTCTTGAATCTTTGCTTATTCGCGAGCTCTACAGTATTTGAT 421
DB 273 AlaCysMetSerLeuSerValTyrIleLeuLeuValLeuTyrIlePheAlaLeuValGly 292
QY 422 CGGCCTACCTTGTGAGAGCTGCTGGATGTTGTTGCTATGGGATGGCCGTAGAT 481
DB 293 ValThrLeuThrLeuProGlyIleAlaAlaPheLeuLeuThrValGlyMetAlaValAsp 312

```



QY 482 GCAATCTCTGTTATTCGAAAGATCCGACGAGAAATTTTATTTGCTCTCAAGTCTTAAA 541  
 Db 313 AlaasnValLeuIlePheGluArgIleGluLeuGlyAsnGlyArgSerIleThr 332  
 QY 542 AAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATCTCAACTTGACT 601  
 Db 333 SerAlaMetIleGlyPheSerAsnAlaLeuArgSerIleMetAspSerAsnIleThr 352  
 QY 602 ACAGTATTTGGCTCAGCACTCTTTCTTCTTCATGATACAGGCGCTTTTAAAGGTTTGGCT 661  
 Db 353 ThrLeuIleAlaGlyLeuValLeuThrPhePheGlySerGlyProValLysGlyPheAla 372  
 QY 662 TTGACATGATTTAGGATTTCTCTTCATGTTTACGGCTCTTTTCATGACATAAATTT 721  
 Db 373 LeuThrLeuLeuIleGlyIleValIleSerMetPheThrAlaIleIleMetThrArgPhe 392  
 QY 722 TTCCTTCATGCTGTGGATGAATAAGACCCCAACATACACAGTTGTCATATGATGAATAAGTTC 781  
 Db 393 PheMetAsnLeuGlyPheAsnMetGlyIleLeuAsnLysProSerMetPheGlyArgIle 412  
 QY 782 GTGGGGATAAGCATGAT 799  
 Db 413 LysGlyGlyArgHisAsn 418

RESULT 21  
 Q9PNK1 PRELIMINARY; PRT; 526 AA.  
 AC Q9PNK1;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Protein-export membrane protein.  
 GN SEC2 OR CJI093C.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OC Campylobacter.  
 OX NCBI\_TaxID=197;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCTC 11168;  
 RX MEDLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies K.M., Feltwell T., Holtroyd S.,  
 RA Jagsels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
 RA Whitehead S., Barrett B.G.;  
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 reveals hypervariable sequences.";  
 RL Nature 403:665-668(2000).  
 DR EMBL; ALJ39077; CAB73348.1; -.  
 DR InterPro; IPR001036; Acriflavin res.  
 DR InterPro; IPR003335; SecD\_SecF.  
 DR Pfam; PF02355; SecD\_SecF-1.  
 DR PRINTS; PR00702; ACRIFLAVINRP.  
 DR TIGRFAMs; TIGR00916; 2A0604s01; 1.  
 DR TIGRFAMs; TIGR01129; secD; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 526 AA; 57251 MW; B2DC392902810CID CRC64;

Alignment Scores:  
 Pred. No.: 1.06e-25 Length: 526  
 Score: 403.50 Matches: 92  
 Percent Similarity: 57.60% Conservativeness: 52  
 Best Local Similarity: 36.80% Mismatches: 97  
 Query Match: 12.18% Indels: 9  
 DB: 16 Gaps: 2

US-09-868-987-1 (1-1864) x Q9PNK1 (1-526)

QY 29 ATCAGCGCACTGCTATGGAATATTTCT-----GCCAAC 64  
 Db 263 IleAsnPheThrLeuAsnAlaGluGlySerLysLysPheAlaAspThrGlyAlaAsn 282

QY 65 CGTGGATGGCGTATGCTGTAGTATTGACGGTTATATGTCAGCAGCCCTATTTTAAAC 124  
 Db 283 ValGlySerArgLeuAlaIleValLeuAspAsnLysValTyrSerAlaProSerIleAsn 302  
 QY 125 GTCCCATTG--AAAATCATCCAGTGTCTCAGGGAAATTTACCCACCGTGAAGTCAGC 181  
 Db 303 GluArgIleGlyGlySerGlyGlnIleSerGlyAlaPheThrGlnGluGluAlaArg 322  
 QY 182 AAATCGCCCTCAGATTTAAATCTGGAGCGATGCTTTTCTTCCCGAGGTTCTCAGTGA 241  
 Db 323 AspValAlaValAlaLeuArgSerGlyAlaLeuAlaProValLysLeuLeuGluGln 342  
 QY 242 GAGACGATCTCTTCTGATCTCTGGAAAAACAATGTACACAGGCATTTCTCAGCATGC 301  
 Db 343 ArgSerIleGlyProSerLeuGlyAlaAspSerIleLysMetSerMetIleAlaLeu 362  
 QY 302 TGTGGCTTGGCAATGCTTATTGTTGATGAGCGTATATTATATAGATTGGAGGGTGCATC 361  
 Db 363 GlyAlaSerValPheIleValValPheMetMetMetTyrTyrGlyValAlaGlyIlePhe 382  
 QY 362 GCTTCGGGAGCTGTTCTCTGAATCTTTTCTTATCTGGGCGAGCTCTACAGTATTTCGAT 421  
 Db 383 AlaAsnIleAlaMetLeuValAsnValValValAlaValAlaMetAlaMetPheGly 402  
 QY 422 GCGCCACTCCTTGTCTCAGACTCGCTGGGATGTTCTTCTGCTATGGGATGGCCGTAGAT 481  
 Db 403 AlaThrLeuThrLeuProGlyMetAlaGlyLeuValLeuThrValGlyMetAlaValAsp 422  
 QY 482 GCAATGTTCTTCTGATTCGAAAGATCCGAGAGGAAATTTTATTGCTCTCAAGTCTTAAA 541  
 Db 423 AlaAsnValIleIleAsnGluArgIleArgIleLeuLeuArgAspGlyValAsnIleArg 442  
 QY 542 AAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCCATTTTGTATCTCAACTTGACT 601  
 Db 443 ValSerIleGluGlnGlyTyrLysAsnAlaMetSerAlaIleIleAspSerAsnIleThr 462  
 QY 602 ACAGTATTTGGCTCAGCACTCTTTTCTTCTGATGATACAGGCGCTTAAAGGTTTGGCT 661  
 Db 463 SerLeuValThrSerValAlaLeuTyrAlaTyrGlyThrGlyAlaValLysGlyPheAla 482  
 QY 662 TTGACATGATTTAGGATTTCTCTTCATGTTTACGGCTCTTTTCATGACATAAATTT 721  
 Db 483 ValThrLeuGlyIleGlyIleValValSerMetIleThrAlaIleLeuGlyThrHisGly 502  
 QY 722 TTCCTTCATGCTGTGGATGAATAAGACCCAA 751  
 Db 503 MetPheAspTyrPheMetGlnArgIleGlu 512

RESULT 22  
 Q8RLP5 PRELIMINARY; PRT; 518 AA.  
 AC Q8RLP5;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Export membrane protein SecD.  
 OS Rickettsia typhi.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=785;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA On'gele B.A., Radulovic S., Azad A.F.;  
 RT "Characterization of the Sec operon of the typhus group Rickettsia.";  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY078352; AAL82790.1; -.  
 SQ SEQUENCE 518 AA; 57125 MW; 6E9E2F2FB717201F CRC64;

Alignment Scores:  
 Pred. No.: 1.91e-25 Length: 518  
 Score: 400.50 Matches: 87  
 Percent Similarity: 58.40% Conservativeness: 52  
 Best Local Similarity: 36.55% Mismatches: 98

Query Match:	12.09%	Indels:	1
DB:	2	Gaps:	1
US-09-868-987-1 (1-1864) x Q8RLP5 (1-518)			
Qy 62 AACCGTGAATGCGCATGCTGTAGTGAATATGCGTATATGCTACAGCCCTTAATTTA 121			
Db 281 AenhrthrGlylsrArgLeuvalAlleValLeuaspHsnlyrLeuenserAlaPromeIle 300			
Qy 122 AAC--GTCCATTGAAAAATCATGACCAGTGTCTTCAGGAAATTTACCCACCGTGAATG 178			
Db 301 AenglyAlaIleIleGlyIYaspclYlleIleThrGlyAsnPhenThrIleGlyserAla 320			
Qy 179 AGCAAACTCGCTCAGATTATAATCTGGACCGCATGCTTTTGTTCGCCAGGTTCACGT 238			
Db 321 AengGluLeuAlaLeuLeuLeuArgValGlyserLeuProthrProLeuYsIleIleGlu 340			
Qy 239 GAAGAGCAGATCTCTTCGATCTTCGGGAAAAAACAATGTACACAGCATTAATCTACGA 298			
Db 341 GluArgserIleGlyProasnLeuClYAlaAsperIleGlyserGlylsrAlaGly 360			
Qy 299 TGCTGTGCTGGCATGCAATGCTATATGTTTGTATGAGCCGTATATTAAGATTGAGCGTC 358			
Db 361 LeuIleGlyPheThrAlaValCysIlePheMetIleLeuSerTYrGlyValIleGlyLeu 380			
Qy 359 ATCCGTTGGGAGCTGTTCTTCTGTAATCTTTTGTCTATCTGGGAGCTCTACAGTATTG 418			
Db 381 PheAlaAsnIleAlaLeuIleLeuAlaLeuLeuYrIleLeuAlaLeuLeuSerLeuPhe 400			
Qy 419 GATGGCCACCTCACCTGTGTCAGAGATCGCTGGATGTTGTTGCTATGGGATGACCGTA 478			
Db 401 GluAlaThrLeuThrLeuProGlyIleAlaGlyIleIleLeuThrMetGlyMetAlaVal 420			
Qy 479 GATGCAATGTCTTGTATGTCGAAAGAATCCGAGAGGAATTTATGTCTCAAGTCTT 538			
Db 421 AspAlaAsnValLeuIleTYrGluArgIleIleGluGluLeuHISlyGlyValSerAsn 440			
Qy 539 AAAAAATCTGTATGAAAAAGATATCCAGGCTTTGGAGCATTTTGTATTTCACTTG 558			
Db 441 LeuTYrAlaIleArgThrGlyPheGluSerAlaPheAlaThrIleIleAspSerAsnLeu 460			
Qy 559 ACTACAGATATGGCTCAGCACTTCTTTCTTCCTAGATACAGGCGCTATTAAGGTTT 658			
Db 461 ThrThrLeuIleValAlaPheAlaLeuYrIlePheGlyValGlyAlaIleIlySGlyPhe 480			
Qy 659 GCTTGAATGATTGATTGAATTTCTCTCAATGTTTACGGCTCTTTCATGATCAATAA 718			
Db 481 AlaValAlaLeuThrIleGlyIleIleIleSerSerMetPheSerAlaIleIleIleThrIly 500			
Qy 719 TTTTCTTCATGCTGTGATGAATADAGCCCAACATACATACAGTTGATGATGATG 772			
Db 501 LeuLeuIleAspIleTYrValIlyTYrPheIlySProIlySLeuGlyLeuLeu 518			
RESULT 23			
Q9CP37 PRELIMINARY; PRT; 616 AA.			
AC 09CP37;			
DT 01-JUN-2001 (TREMBLrel. 17, Created)			
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE SecD.			
GN SEC D OR PM0227.			
OS Pasteurella multocida.			
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;			
CC Pasteurella			
OK NCBI_TaxID=747;			
OK [1]			
RN SEQUENCE FROM N.A.			
RP STRAIN=PM70;			
RX MEDLINE=21145866; PubMed=11248100;			
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitam T.S., Kapur V.,			
RT "Complete genomic sequence of Pasteurella multocida Pm70.",			
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).			

DR	EMBL; AEO06057; AAK02311.1; -	DR	InterPro; IPR001036; Acrlfvln_res.
DR	InterPro; IPR003335; Secd_SecF.	DR	Pfam; PF00355; Secd_SecF; 1.
DR	PRINTS; PR00702; ACRIPLAVINRP.	DR	TIGRFAMs; TIGR00916; 2A0604901; 1.
DR	TIGRFAMs; TIGR01129; secod; 1.	DR	TIGRFAMs; TIGR01129; secod; 1.
KW	Complete proteome.	SEQUENCE	616 AA; 67401 MW; 9BBE98497DFB8403E CRC64;
US-09-866-987-1 (1-1864) x Q9CP37 (1-616)			
QY	137 AATCATGCCAGTGTCTCAGGAAATT-----ACC	166	
Db	395 AsnValaAatrrrlliegInglYarPheglYserGInPhegnllrthrglYleasPser	414	
QY	167 CACCGTAGAGGAGCAACTGCGCTCAGATTAAATCGAGGANGCTTTTGTTCCC	226	
Db	415 ProAlaGluAlaAleuAsnleuAlaValleuLeuYrserglYalAleuileAlaProile	434	
QY	227 GAGGTCTCAGTGAAGAGACCATCTCTTCGATCTTGGAGAAAACAAATGACAAAGC	286	
Db	435 GlnleValaIglnlUarGThrllleGlyProserleuGlYalGluenValaGlnGlnly	454	
QY	287 ATTATCTCAGCATGCTGTGGCTTTGGCAATGCTTATTTGTTGAGCGTAATATATAGA	346	
Db	455 LeuGlnAlaSerPheTrpGlyLeuLeuAlaVallellePheMetThrleuYrTYrArg	474	
QY	347 TTTGGAGCGATCAGTCGCTGGAGCGTCTTCTGAAATCTTTGCTTATCTGGAGCGT	406	
Db	475 LysPheGlyLeuIleAlaAsnIleAlaLeuIleAlaAsnIleValleuAlaGlyLeu	494	
QY	407 CTACAGATTTTG--GATGCGCCACTCACCTTGCAGAGCTCGTGGATTTGTTGCT	463	
Db	495 MetSerleuLeuProglYAlaThrleuSerMetProglYleAlaGlylleValleuSer	514	
QY	464 ATGGGATGCGCGTAGATGCAAATGTTCTTGATTCGAAAGAAATCGAGAGAAATTTTA	523	
Db	515 ValGlyMetSerIleAspAlaAsnValleuIlePheGlnUarGlyleuGlnGlnIleArg	534	
QY	524 TTGTCTCAAAAGCTTAAAAAATCTGTGMAAAAGGATATACCAAGGTTTGGAGCATT	583	
Db	535 AsnGlyUarProIleGlnAlaIleIleGlnIlyYrserglYAlaPheTrpMetIle	554	
QY	584 TTGATTTCTACTTGACTACACAGTATTTGGCCACAGCACTTCTTTCTTCAGATACAGG	643	
Db	555 PheAlaAlaAsnleuThrThrIleleuThrSerValleuYrAlaValaGlyAlaGly	574	
QY	644 CCTATTAAAGGTTTGCTTTGACATTGATTTAGAAATTTCTCTTGAATGTTACGGCT	703	
Db	575 ProValYsglyPheAlaIleThrleuAlaIleuGlYAlaGlyIleSerMetPheThrAla	594	
QY	704 CTTTCACTGACTAA 718		
Db	595 IleValGlyThrArg 599		
RESULT 24 .			
ID	Q92H77	PRELIMINARY;	PRT; 518 AA.
AC	Q92H77		
DT	01-DEC-2001 (TReMBLrel. 19, Created)		
DT	01-DEC-2001 (TReMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (TReMBLrel. 21, Last annotation update)		
DE	Protein-export membrane protein secod.		
GN	SECOD OR RC0894.		

```

OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiaceae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MALISH 7; PubMed=11557893;
RX MEDLINE=21442074;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
DR EMBL; AE008643; AL003432.1; -.
DR InterPro; IPR003335; SecD SecF.
DR PFAM; PF02355; SecD SecF; 1.
DR TIGRFAMs; TIGR00916; 2A0604s01; 1.
DR TIGRFAMs; TIGR01129; secD; 1.
KW Complete proteome.
SQ SEQUENCE 518 AA; 56516 MW; 4487A46517B65CB4 CRC64;

Alignment Scores:
Pred. No.: 5,06e-25 Length: 518
Score: 395.50 Matches: 86
Percent Similarity: 59.03% Conservative: 48
Best Local Similarity: 37.89% Mismatches: 92
Query Match: 11.94% Indels: 1
DB: 16 Gaps: 1

US-09-868-987-1 (1-1864) x Q92H77 (1-518)
QY 62 AACGTCGATGGCGTATGCTGTAGTATGATGACGGTATATGTCAGCAGCCCTATTTTA 121
Db 281 AsnileGlyLysArgLeuAlaIleValLeuAspAsnLysLeuLeuSerAlaProThrIle 300
QY 122 AACGTCGCATGAAAATCATGCCAGT---GTCTCAGGGAATTTACCCAGCGTGAAGTG 178
Db 301 AsnGlyAlaIleMetGlySerGlyIleThrGlyAsnPheThrValGluSerAla 320
QY 179 AGCAAACTCGCTCAGATTTAAATCTGGACGATGCTTTTGTTCGCGAGTTCTCAGT 238
Db 321 AsnGluLeuAlaLeuLeuLeuArgAlaGlySerLeuProAlaProLeuLysIleIleGlu 340
QY 239 GAAGAGACGATCTCTCTGATCTTTGGGAAAAACAATGTACACAGGCAATTCATCTCAGCA 298
Db 341 GluArgSerIleGlyProAsnLeuGlyAlaAspSerIleGluSerGlyLysLysAlaGly 360
QY 299 TGCTGTGCTGGCAATGCTATTGTTTGTATGACGCTATATATAGATTGGAGGCGTC 358
Db 361 LeuIleGlyPheIleAlaValCysIlePheMetValTrpSerTrpGlyValLeuGlyLeu 380
QY 359 ATCGCTTCGGAGCTGTTCTTCTGAATCTTTTGTCTTCTGCGGAGCTCTACAGTATTG 418
Db 381 PheAlaAsnIleAlaLeuSerLeuAlaLeuLeuTyriIleLeuAlaLeuSerLeuPhe 400
QY 419 GATCGCCACTCACCTGTCTCAGGACTCGCTGGGATTTGTTCTTGTATGGGATGCGCGTA 478
Db 401 GlnAlaThrLeuThrLeuProGlyIleAlaGlyIleIleLeuThrMetGlyMetAlaVal 420
QY 479 GATCAATGTTCTTGTATTCGAAGAATCCGAGAGAAATTTTATTGCTCAAGAGCTT 538
Db 421 AspAlaAsnValLeuIleTyriGluArgIleLysGluGluLeuHisLysGlyValSerThr 440
QY 539 AAAAAATCTGAGAAAGGATATACCAAGCTTTTGGAGCCATTTTGTCTTAACTTG 598
Db 441 LeuTyriAlaIleArgThrGlyPheGluSerAlaPheAlaThrIleLeuAspAlaAsnLeu 460
QY 599 ACTACGATTTGGCTCAGCACTCTTTTCTCTCTAGATACAGGCGCTATTAAAGGGTTT 658
Db 461 ThrThrLeuIleValAlaPheLeuLeuTyriIlePheGlyValGlyAlaIleLysGlyPhe 480
QY 659 GCTTTGACATTTAGGAATTTCTTCTCAATGTTTACGGCTCTTTTCATGACTAAA 718

```

---

```

Db 481 AlaValAlaLeuThrIleGlyIleIleSerSerMetPheSerAlaIleIleIleThrLys 500
QY 719 TTTTCTTTCATGCTGTGGATG 739
Db 501 LeuLeuIleAspIleTrpVal 507
RESULT 25
Q97GT7
ID Q97GT7 PRELIMINARY; PRT; 417 AA.
AC Q97GT7;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Preprotein translocase subunits SecD.
GN CAC2278.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillia/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AB007729; AAK80235.1; -.
DR InterPro; IPR003335; SecD SecF.
DR PFAM; PF02355; SecD SecF; 1.
DR TIGRFAMs; TIGR00916; 2A0604s01; 1.
DR TIGRFAMs; TIGR01129; secD; 1.
KW Complete proteome.
SQ SEQUENCE 417 AA; 44109 MW; 8B073289DCF191AD CRC64;

Alignment Scores:
Pred. No.: 1.19e-24 Length: 417
Score: 391.00 Matches: 77
Percent Similarity: 61.86% Conservative: 69
Best Local Similarity: 32.63% Mismatches: 82
Query Match: 11.81% Indels: 8
DB: 16 Gaps: 3

US-09-868-987-1 (1-1864) x Q97GT7 (1-417)
QY 68 GGATGGCGTATGGCTGATGATGACGGTTATATGTCAGCAGCCCTATTTTA---AAC 124
Db 173 GlyGlnLysIleSerIleTyriMetAspThrAspLeuLeuSerAspProThrValAspSer 192
QY 125 GTCCCATGAAAATCATGCCAGTCTCAGGNAATTTACCCACCGTGAAGTGAAGCAA 184
Db 193 ValIleThrGlyGlyAsnAlaGlnIleThrGlyAsnLysSerIleGlnGluAlaGlnArg 212
QY 185 CTCGCTTCAGATTTAAATCTGGAGCATGCTCTTTTCTCCGAGGTTCTCAGTGAAGAG 244
Db 213 IleAlaAsnIleLeuAsnSerGlyAlaLeuProValThrLeuLysValGlnSerLys 232
QY 245 ACGATCTCTTCTGATCTTGGGAAAAACAATGTACACAGGCATTTATCTCAGCATGCTGT 304
Db 233 ThrValGlyAlaSerLeuGlyAlaSerAlaLeuProAsnSerIleLeuAlaGlyAlaVal 252
QY 305 GGCTTGGCAATGCTTATTGTTTGTATGAGCGGTATATTATAGATTTGGAGGCGTCATCGCT 364
Db 253 AlaIleCysIleIlePheLeuPheMetIleLeuValTyriArgIleProGlyLeuMetAla 272
QY 365 TCGGAGCTGCTCTTCTGATCTTTTGTCTTATCTGGGACGCTCTACAGTATTTCGATGCG 424
Db 273 AspIleAlaValLeuPheThrValLeuValLeuGlyAlaPheValAlaValLysVal 292
QY 425 CCACTACCTTGTTCAGGACTCGCTGGGATTTCTTCTGCTATGGGATGGCGGTAGATGCA 484

```

```

Db 293 ThrLeuThrLeuSerGlyIleAlaGlyLeuLeuLeuThrIleGlyMetAlaValAlaSpA 312
Qy 485 AATGCTTCTGATTCGAAAATAATCCGAGAGAATTTTATGTCCTCAAGCTTAAAAA 544
Db 313 AsnValIleuIlePheGlyAlaGlyPheGlySerGlyIleGlyValThrValArgSer 332
Qy 545 TCTGTAGAAAAAGATATACCAAGCTTTTGAGCCATTTTGATTTGATCTGATCA 604
Db 333 AlaPheAspAlaGlyPheIleArgAlaMetSerSerIleIleAspSerAsnValThrThr 352
Qy 605 GTATGGCCCTCAGACACTTTCTTTCTTCTAGATACAGGCGCTATTAAGGCTTTGCTTG 664
Db 353 IleIleSerGlyValIleLeuThrAlaLeuGlySerGlyValGlyValGlyPheAlaLeu 372
Qy 665 ACATTGATTTTAGAATTTCTCTTCAATGTTTACGCGCTTTTCAATGATCAATTTTTC 724
Db 373 ThrLeuValIleGlyValIleValIleuIleSerMetValThrAlaIleIleValThrArgLeu 392
Qy 725 TTCATGCTGTGG-----ATGAAATAGACCCACAT 754
Db 393 LeuMet---TTPAlaIleAlaPheMetGlyLeuIleAsnGlyAlaSerHis 407

RESULT 26
ID 09A6U2 PRELIMINARY; PRT; 532 AA.
AC 09A6U2;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Protein-export membrane protein Secd.
GN CC1991.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
NCBI_Taxid=155892;
RX NCBI [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RC MEDLINE=21173698; PubMed=11259647;
RA Nielsen W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eissen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A.S., Stephens C., Phadke N.D., Ely B.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uetereback T., Tran K., Wolf A., Vamathevan J., Ernoileva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005872; AKK23966.1; -.
DR TIGR; CC1991; -.
DR InterPro; IPR003335; Secd_SecF.
DR Pfam; PF02355; Secd_SecF.1.
DR TIGRFAMs; TIGR00916f_2A0604s01; 1.
DR TIGRFAMs; TIGR01129; secD; 1.
KW Complete proteome.
SQ SEQUENCE 532 AA; 56792 MW; 80A85984798F64C7 CRC64;

Alignment Scores:
Pred. No.: 1.22e-24 Length: 532
Score: 391.00 Matches: 85
Percent Similarity: 60.34% Conservative: 58
Best Local Similarity: 35.86% Mismatches: 90
Query Match: 11.81% Indels: 4
Gaps: 3

US-09-868-987-1 (1-1864) x 09A6U2 (1-532)
Qy 62 AACCGTGAATGCGGTATGCTGATGATGAAGGTATATATGCTCAGACCCCTATTTTA 121
Db 295 AsnValIleGlyValArgPheAlaIleValIleuAspGlyValGlyIleIleSerAlaIleProThrIle 314
Qy 122 AAC---GTCCATTGAAAAATATATGACGACTGTCTAGAGGAAATTTACCCACCGTAAAGTG 178

```

```

Db 315 AsnGlyAlaIleLeuGlyGlySerGlyIleIleThrGlySerPheThrAlaGlySerAla 334
Qy 179 AGCAACTCGCTCAGATTTTAAATCTGAGCCGATGCTTTGTTCCGAGCTTCTAGT 238
Db 335 SerAspLeuAlaLeuLeuLeuArgSerGlyAlaLeuProAlaProLeuValGluGln 354
Qy 239 GAAGAGCAATCTCTGATCTTGAATCTTGAGAAAAACAATGTACACAGGCAATATCTAGCA 298
Db 355 GlnAsnThrValGlyAlaGlyLeuGlyAlaAspAlaValAlaGlyAlaIleSerThr 374
Qy 299 TGCTGTGGCTTGGCAATCTTATGTTGTTGATGAGCCATATATTATTT-----AGATTGGA 352
Db 375 LeuValAlaPheIleThrIleValAlaPheMetIleLeuSerGlyValLeuLeuPheGly 394
Qy 353 GGGCTATCGCTTGGAGAGCTGTCTTCTGAATCTTTGCTATCTGGCAGCTCTACAG 412
Db 395 Gly---IleSerValIleAlaLeuIleIleAsnGlyMetLeuIleValAlaAlaMetSer 413
Qy 413 TATTGGATGGCCACATCCACTTGTCTGACAGCTGCTGGAGTTGTTCTGATAGGGAGTG 472
Db 414 LeuThrGlnAlaThrLeuThrLeuProGlyIleAlaGlyLeuIleLeuThrLeuAlaVal 433
Qy 473 GCCGTAGATGCAATGTTCTTGTATTCGAAAGATCCGAGAGAAATTTTATGTCACAA 532
Db 434 AlaValAspAlaAsnValIleuIleGlyValArgMetArgAspGlyAlaGlyLys 453
Qy 533 AGTCTTAAAAATCTGAGAAAAAGATATATCAAGGCTTTTGAGCCATTTTGAATCT 592
Db 454 SerProIleLeuAlaAlaAspAlaGlyPheSerArgAlaMetThrThrIleIleAspAla 473
Qy 593 AACTGACTACAGTATGGCTGACACTCTTTCTTCCAGATATACAGGCGCATTA 652
Db 474 AsnValThrThrLeuValAlaAlaGlyIleMetPheAlaPheGlyAlaGlyProValArg 493
Qy 653 GGGTTTCTTGAATGATTTTGAATTTTCTTCAATGTTTACGCTTTTTCATG 712
Db 494 GlyPheAlaThrPheThrLeuSerIleGlyValPheHisSerValPheThrAlaValLeuVal 513
Qy 713 ACTAAATTTTCTTCATGCTGTGTGATGAATATAGACCCACATACACAGTTG 763
Db 514 SerGlnLeuLeuIleGlyTyrThrPheArgAlaAlaArgProGlyLysLeu 530

RESULT 27
ID 09PNK2 PRELIMINARY; PRT; 323 AA.
AC 09PNK2;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Protein-export membrane protein.
GN SecF OR CUI092C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
NCBI_Taxid=197;
RX NCBI [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RC MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kerley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
RA Jørgensen K., Kariyasek A.V., Moule S., Pallen M.J., Peum C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139077; CA873347.1; -.
DR InterPro; IPR003335; Secd_SecF.
DR Pfam; PF02355; Secd_SecF.1.
DR TIGRFAMs; TIGR00916f_2A0604s01; 1.
DR TIGRFAMs; TIGR00966; 3a0501s07; 1.

```

```
KW Complete proteome.
SQ SEQUENCE 323 AA; 36112 MW; 23589FD95B5BFD02 CRC64;

Alignment Scores:
Pred. No.: 1,41e-24 Length: 323
Score: 390.00 Matches: 98
Percent Similarity: 53.09% Conservative: 74
Best Local Similarity: 30.25% Mismatches: 96
Query Match: 11.78% Indels: 56
DB: 16 Gaps: 9

US-09-868-987-1 (1-1864) x Q9PNK2 (1-323)
QY 794 CATGATTTCTTGAGAGATGCAAAACCTTTGGCGTGTTC----- 835
Db 10 TyrAspPheMetArg-----MetArgPheAlaAlaIleSerLeuSerPheIleLeuPhe 27
QY 836 ---GGAAGTGTTCCTTTAGTTGCGTCTCTCGGGTTGGAGCTGGAATTCGTT 892
Db 28 PheGlySerIleTyrLeuLeu-----TrpAspArgGly 38
QY 893 TTG-----GGAATGGATTTAAAGAGGGTATGCTTTTACCTTTAATCCAAAGAGCAT 946
Db 39 LeuGlnTyrGlyIleAspPheSerGlyGlyThrLeuValGlnLeu-----LysTyrGlu 56
QY 947 GGCATCAGCGATGTTGCTCAAAATCGGTGGCAAAAGTTGTGCATAACTACAGGAAGCTGGT 1006
Db 57 AsnAlaAlaProIleThrGlnIleArg-----GluIleLeuGluAsnGln 71
QY 1007 CTTCTCTAGAGACTCCGTATTCACAACTTTGATCTTCAGAAAGATCAAAATCTAT 1066
Db 72 GlyThrPheGlnAsnLeuSerValThrGluPheGlySerAsnGluValThrIleArg 91
QY 1067 TTT-----AGTGATAAGCTTTAAGCTATATAAGCAGATACGAGCTCTCTCTA 1117
Db 92 PheLeuGlySerAsnAspAsnValSerAsnAspIleGlyGluHisIleSerThrLeuLeu 111
QY 1118 AAATTAACGATCATGAGCTGCGGTATTGTGGGATTTGTGTGAGAAACAGCGCTAGATTT 1177
Db 112 LysAspThr----- 114
QY 1178 CTCTACGGAACCTTAACAGAAACGCAAAATTTGGTCAAAAGTAAAGCAGCAAACTATCG 1237
Db 115 -----GlyLysPheGluValArgAlaAspValValGlyProLysValGly 130
QY 1238 AAGAAATGCTTATCAGCGCACCATCGGCTTTTAGGAGCTTTGGCAATCATCTTGCTC 1297
Db 131 AspGluLeuArgAsnLysGlyLeuMetAlaIleAlaValSerLeuIleAlaIleLeu 150
QY 1298 TATGTGAGTTTGGCTTTGAATGGCAATATGCTTTTCAAGTGGCTGATGCGCTTTAATTCAT 1357
Db 151 TyrIleAlaLeuArgPheGluTyrArgPheAlaLeuAlaIleIleSerGluIleHis 170
QY 1358 GACCTTTGGCTACTGTGCACTGTTGTTATAGCACATTTCTTTTGAAGAAATTCAA 1417
Db 171 AspValValIleThrLeuGlyAlaIle-----SerLeuPheLysIleAsp 185
QY 1418 ATAGATTGCAAGCATTTGGTCTTTAATGACTGTATTGGGTATTCATTAAACAATACT 1477
Db 186 ValAsnLeuAspThrLeuAlaAlaValLeuThrValLeuGlyTyrSerLeuAsnAspThr 205
QY 1478 TTGATCATTTTGTGATTCGTGAAGATCGCCAAAGCGAACCTGTTTACCCCTATGCA 1537
Db 206 IleIleIlePheAspArgIleArgGluGlyIleLysThrSerLysLysThrGluLeuAla 225
QY 1538 GTTTTAGTTAATGATGCCCTTCAAAAGAGTTTCAGCCGCAAGTAATGAAACAGCTACA 1597
Db 226 ProIleIleAsnGluSerValSerAlaThrLeuSerArgThrValLeuThrSerGlyLeu 245
QY 1598 ACTCTATCAGTTTGTAAATCTTTGTTTATAGCGGCTCTCTGCTTTTAAATTTGCA 1657
Db 246 ThrLeuAlaThrValIleLeuTyrPhePheGlyGlyGluMetIleGlnGlyPheSer 265

QY 1658 TTTATTATAGACCATAGGATTCCTTAGGAACCTTTATCGTCTCTTATATATGACCACT 1717
Db 266 LeuAlaLeuIleValGlyIleIleAlaGlyThrLeuSerSerIlePheValAlaSerPro 285
QY 1718 CTGTTGTTGTTT 1729
Db 286 ThrLeuLeuTyrPhe 289

RESULT 28
Q8ZC36 PRELIMINARY; PRT; 322 AA.
AC Q8ZC36;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-export membrane protein SecF.
GN SECFO YPOJ188.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
ON NCBI_TaxID=632;
RX [1]
SEQUENCE FROM N.A.
STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feitwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skellerton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL; AJ414155; CAC92423.1; -
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF; 1.
DR TIGRFAMs; TIGR00916; 2A0604801; 1.
DR TIGRFAMs; TIGR00966; 3a0501s07; 1.
KW Complete proteome.
SQ SEQUENCE 322 AA; 35178 MW; 2029AEB45F281ADC CRC64;

Alignment Scores:
Pred. No.: 4,13e-24 Length: 322
Score: 384.50 Matches: 104
Percent Similarity: 52.88% Conservative: 61
Best Local Similarity: 33.33% Mismatches: 114
Query Match: 11.61% Indels: 33
DB: 16 Gaps: 10

US-09-868-987-1 (1-1864) x Q8ZC36 (1-322)
QY 794 CATGATTTCTTGAGAGATGCAAAACCTTTGGCTGTTTCTGGAAGTGTTCCTTTTA 853
Db 17 TyrAspPheMetArgTyrAspTyrValAlaPheGlyValSerLeuLeuValAla 36
QY 854 GGTTCGCTGCTCTCGGGTTTGGAGCTGCAATTCCTTTTGGCAATGGATTAAAGGA 913
Db 37 SerIleValValMetSerThrLysGlyPheAsn-----TrpGlyLeuAspPheThrGly 54
QY 914 GGGTATGCTTTTACCTTTAATCAAAAGAGCATGGCATCAGCGATGTTCCTCAATGCGT 973
Db 55 GlyThrValIleGluIleAsnLeuGluAsnPro-----AlaAspLeuAspGlnLeuArg 72
QY 974 GGCAAACTGTGCATAAATCTACAGGAAGCTGTTCTTCTCTAGAGACTTCGCTATTCAA 1033
Db 73 -----AspThrLeuGlnAsnAlaGlyPheGluSerPro-----IleLeuGln 86
QY 1034 ACATTTGGATCTTCAGAAAGATCAAAATCTATTTTAGTGATAAAGCTTTAAGCTATCT 1093
Db 87 AsnPheGlySerArgAspValMetValArgMetProAlaThrGlyThrAlaGly 106
QY 1094 AAGCAGATACGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1153
```



Db 393 LeuHiaAlaLeuLeuGluMetAspLeuThrArgAenValLysPheTyr 408

RESULT 31

Q9KTY7

ID Q9KTY7 PRELIMINARY; PRT; 617 AA.

AC Q9KTY7;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Protein-export membrane protein SecD.

GN VC0743.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI\_TaxID=666;

ON [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EL TOR N16961 / SEROTYPE O1;

RX MDLINE=20406833; PubMed=10952301;

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,

RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,

RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Dragoi I., Sellers P.

RA Ramolaeva M.D., Vamathevan J., Bass S., Qin H., Racholai I.,

RA McDonald L., Uitterback T., Fleischmann R.D., Niernan W.C., White O.,

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,

RA Fraser C.M.;

RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio

RT cholerae";

RL Nature 406:477-483(2000).

DR ENBL; A0004160; AAF93908.1; --.

DR TIGR; VC0743; --.

DR InterPro; IPR001036; Acrlflvln res.

DR InterPro; IPR003335; SecD SecF.

DR Pfam; PF02355; SecD SecF.1.

DR PRINTS; PR00702; ACRIFLVINRP.

DR TIGRFAMS; TIGR00916; 2A0604s01; 1.

DR TIGRFAMS; TIGR01129; secD; 1.

KW Complete proteome.

SQ SEQUENCE 617 AA; 66841 MW; 69B93D75E0DE014C CRC64;

Alignment Scores:

Pred. No.: 2,318-23 Length: 617

Score: 376.00 Matches: 91

Percent Similarity: 57.32% Conservative: 46

Best Local Similarity: 38.08% Mismatches: 82

Query Match: 11.35% Indels: 20

DB: 16 Gaps: 3

US-09-868-987-1 (1-1864) x Q9KTY7 (1-617)

Qy 62 AACCGTGTAGCGGTATGCGTGTAGTGTACGGTTATATGTCAGC-----AGC 112

Db 362 AsnIleGlyLysLeuMetAlaThrValPheAlaGluTyrLysAspSerGlyLysArgSer 381

Qy 113 CCTATTTAAACGTCCTCATGAAAAATCATGCCAGTGTCTCAGGGAAATTTACCCACCGT 172

Db 382 ProGluGlyLysValIleLeuThrLysHisGluValIleAenGlnAlaThrIleGln 401

Qy 173 -----CAAGTGAGCANA 184

Db 402 SerAlaLeuGlyArgAsnPheArgIleThrGlyIleAspSerProAlaGluAlaHisAsn 421

Qy 185 CTCGCCCTCAGATTAAAACTCTGGAGCGATGTCCTTTGTTCCCGAGGTCTCTCAGTGAAGAG 244

Db 422 LeuAlaLeuLeuLeuArgAlaGlyAlaLeuIleAlaProIleSerIleValGluGluArg 441

Qy 245 ACGATCTCTTCGATCTGGGAAAAAACAATGTACAAAGGCATTATCTCAGCATGCTGT 304

Db 442 ThrIleGlyProSerMetGlyGlnGinAenIleAspMetGlyIleGlnAlaCysIleTyr 461

Qy 305 GCGTTGCAATGCTTATGTTGTTGATGACGGTATATTATGATTTGGAGCGCTCATCGCT 364

Db 462 GlyMetValAlaValMetLeuPheThrValLeuTyrArgLysPheGlyMetIleAla 481



```

QY 365 TCGGAGCTGTTCTTCTGATCTTTTGTATCTGGGAGCGCTTACGATATTG---GAT 421
Db 482 Asn1leAlaLeuMetAlaAsnleuValleu1le1leGlyValMeSerMetileProGly 501
QY 422 GCGGCACCTACCTTGTCTGACGACGCTGCGGATTTGTTCTGTATGGGANTGGCGGTAGAT 481
Db 502 AlaThrMetThrLeuProGly1le1leGly1leValleuThrValGlyMetAlaValasp 521
QY 482 GCAATGTTCTTGTATTTGGAAGAATCCGAGAGCAATTTTATTTGTTCTCAAAAGTTTAA 541
Db 522 AlaAsnValleu1lePheGly1leArg1leArg1leuArg1leuArg1leuArg1leuArg1leu 541
QY 542 AAATCTGAGAAAAGATATACCAAGCTTTTGAGCGCATTTTGTATTTGATCTGATCTGACT 601
Db 542 Gln1le1leHisGlnGly1leAlaAsn1lePheSerThr1leAlaAspAlaAsn1leThr 561
QY 602 ACAGTATGCGCTGACGACCTCTTTCTTCTGATACAGGCGCTTATTAAGGTTTGTCT 661
Db 562 Thrleu1leThrAla1le1le1leuPheAlaValGlyThrGlyAla1le1leGlyPheAla 581
QY 662 TTGACATTTGATTTAGAAATTTTCTCTTCAATGTTTACGGCTCTTTTGCATGACTTAA 718
Db 582 ValThrleuSer1leGly1le1leuThrSerMetPheThrAla1leValGlyThrArg 600

RESULT 32
Q8XVW2
ID Q8XVW2 PRELIMINARY; PRT; 626 AA.
AC Q8XVW2;
DT 01-MAR-2002 (TReMBLrel. 20; Created)
DT 01-MAR-2002 (TReMBLrel. 20; Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21; Last annotation update)
DE Probable export membrane protein SEC2 transmembrane.
GN SEC2 OR RSC2715 OR R50081.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Sclenoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Catalicio L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunne S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siquier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.,
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AF646071; CADI6422.1; -
DR InterPro; IPR003335; Sec2_SecF.
DR Pfam; PF02355; Sec2_SecF.1.
DR TIGRFAMs; TIGR00916; 2A0504s01; 1.
DR TIGRFAMs; TIGR01129; sec2; 1.
KW Complete proteome.
SQ SEQUENCE 626 AA; 66810 MW; DCC2F59A5FAC2F7B CRC64;

```

```

Alignment Scores:
Pred. No.: 4,15e-23 Length: 626
Score: 373.00 Matches: 89
Percent Similarity: 56.12% Conservative: 44
Best Local Similarity: 35.44% Mismatches: 100
Query Match: 11.26% Indels: 4
DB: 16 Gaps: 1
US-09-868-987-1 (1-1864) x Q8XVW2 (1-626)
QY 62 AACGCTGATGGCGTATAGCTGATGATTT-----GACGTTATATGCTACG 109
Db 371 Asn1leGly1leArgMetGly1leValleuPheGly1leGly1leValleuThr 390
QY 110 AGCCATTATTAACGTCCTCCATTTGAAAATATCATGCGACATGCTCAGGAAATTTACCAAC 169

```

```

Db 391 ValAlaThr1leGlnSerGly1leuGlySerArgPheGln1leThrGly1leGlySerVal 410
QY 170 CGTGAATGTGCAAACTCGCGCTCAGATTTAAATCTGGAGCGATGCTTTGTTCCGAG 229
Db 411 GlnSerAlaSerAspLeuAlaLeuLeuArg1leGlySerLeuAla1lePromeGlu 430
QY 230 GTTCTGATGAGAGAGCATCTCTTGTGATCTTGGGAAAAAACAATGATACACAGGCTT 269
Db 431 Ile1leGln1leuArgThr1leGlyProSer1leuGly1leAspAsn1le1leGlyPhe 450
QY 290 ATTCACAGATCGTGTGCTTGGCAATGCTTATTTGTTGATGAGCGATATATGATTT 349
Db 451 AsnSerAla1leArg1lePheAla1le1leSerValPheMetValleu1leu1leu1leu 470
QY 350 GAGGCGTCAATCGCTTGGGAGCTGTTCTTGTGATCTTTGCTTATCTGGGAGCTCTA 409
Db 471 PheGly1leAlaPheSerVal1leAla1leuGly1leAla1leuValPheLeu1leAla1leu 490
QY 410 CAGTATTTGATGGCGCACCTGCTGAGAGCTGCGTGGATTTGTTGCTGATGGG 469
Db 491 SerMetLeuGln1leAlaThr1leuThr1leuProGly1leAla1leAla1leAla1leuGly 510
QY 470 ATGGCCGTAGATGCAATGTTCTTGTATTCGAAAGAATCCGAGAGCAATTTTATTTGCT 529
Db 511 MetAla1leAspAlaValleu1le1leAsnGluArgValArg1leu1leuArgAsnGly 530
QY 530 CAAGCTTTAAAAATGTGAGAAAAGATATATCAAGGCTTTTGAGGCAATTTTGTAT 589
Db 531 AlaSerProGlnMetAla1leAla1leAlaGlyPheGluArg1leAla1leThrAla1leThr 550
QY 590 TCTACCTGATCAGATTTAGGCTCGACGACTCTTTCTTCTGATATACAGGCGCTATT 649
Db 551 SerSerValThr1leThr1leu1leAlaGlyLeuAla1leu1leuAlaPheGlySerGlyProVal 570
QY 650 AAAGGTTTCTTGTGACATTTTATTTAGAAATTTCTTCAATGTTTACGGCTTTTC 709
Db 571 ArgGlyPheAlaValAlaValHisGlySerGly1le1leuThrSerMetPheSerAlaValPhe 590
QY 710 ATGACTAAATTTTCTTCATGCTGTGATGAATATGACCAATACACG 760
Db 591 PheAsnArgSerLeuValAsnLeu1leu1leu1leu1leu1leu1leu1leu1leu1leu 607

RESULT 33
Q9HX12
ID Q9HX12 PRELIMINARY; PRT; 306 AA.
AC Q9HX12;
DT 01-MAR-2001 (TReMBLrel. 16; Created)
DT 01-MAR-2001 (TReMBLrel. 16; Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21; Last annotation update)
DE Secretion protein SecF.
GN SEC2 OR PA3820.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan X.,
RA Brody K.L., Coulter S.N., Folger K.R., Kas A., Larbig I.T.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olsson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004799; AAG07207.1; -
DR InterPro; IPR003335; SecF_SecF.
DR Pfam; PF02355; Sec2_SecF.1.
DR TIGRFAMs; TIGR00916; 2A0604s01; 1.
DR TIGRFAMs; TIGR00966; 3a0501s07; 1.

```



```
KW Complete proteome.
SQ SEQUENCE 306 AA; 33041 MW; 7C078C8409DF9BA2 CRC64;

Alignment Scores:
Pred. No.: 7,67e-23 Length: 306
Score: 369.50 Matches: 91
Percent Similarity: 51.69% Conservative: 77
Best Local Similarity: 28.00% Mismatches: 102
Query Match: 11.16% Indels: 55
DB: 16 Gaps: 10

US-09-868-987-1 (1-1864) x Q9HX12 (1-306)
QY 779 TTCGTGGGTAAGCATGTTCTTGAGAGATGCAAAAACCTTTGGGCTGTTCTTCTGGA 838
Db 10 PheMetGlyIleArgAsn-
QY 839 AGTGTTTCTTCTAGGTTGGCTCTCGGTTTGGAGCTGGAATCCGTT-----892
Db 19 AlaValThrLeuIle---LeuThrValIleAlaLeuGlySerIlePheThrLysGlyIle 37
QY 893 ---TGGGAATGGAATTTAAAGAGGGTATCGCTTTTACCTTTAATCCAAAAGAGCATGGC 949
Db 38 AsnPheGlyLeuAspPheThrGlyThrLeuIleGluLeuThrTyrGluGlnPro---56
QY 950 ATCAGCATGTTGCTCAATCGTGGCAAGTGTGTCATAAACTACAGGAAGCTGGTCTT 1009
Db 57 ---AlaAspLeuGlyLysValArgGlyGlnLeuVal-----GlyAla 69
QY 1010 TCTTCTAGAGACTCCGTTATCAACATTTTGGATCTTCAGAAAGATCAAAATCTATTT 1069
Db 70 GlyTyrGluAspAlaValAlaGlnSerPheGlyAspAlaArgAspValLeuValArgMet 89
QY 1070 AGTGATAAGCTTTAAGCTATATAAGCAGATACGAGCTCTCTCTAAAATTAACGATC 1129
Db 90 ProSerGluAspProGluLeuGlyLysValAlaThrAlaLeuGlnGlnAlaAsp---108
QY 1130 ATGAGCTGGCGTTATGTGGGATTGTGTCTAGAAACAGCGCTAGATTCTCTACGGAAAC 1189
Db 109 -----AlaGly 110
QY 1190 TCTAAACGAAACCAAAATTTTGGTCAAGGTAAAGCAGCAAACTATCGAAGAAATCGGT 1249
Db 111 AsnProAlaAsnLeuLysArgValGluTyrValGlyProGlnValGlyGluLeuArg 130
QY 1250 TATCAGCGCATCGGCTTTTAGGAGCTTTGGCAATCATCTGCTCTATGTGAGTTG 1309
Db 131 AspGlnGlyLeuGlyMetLeuLeuAlaLeuGlyGlyIleLeuLeuTyrValGlyPhe 150
QY 1310 CGTTTGAATGGCAATATGCTTTTCAGTCCGATATGCGCTTTAATTCATGACCTTTGGCT 1369
Db 151 ArgPheGlnTrpLysPheAlaLeuGlyAlaIleLeuSerLeuValHisAspAlaIleIle 170
QY 1370 ACCTGTGCGCTGTTTATAGCATTCTTTTGAAGAAATCAATAGATTGCAA 1429
Db 171 ValMetGlyValLeu-----SerPhePhe-----GlnValThrPheAspLeuThr 185
QY 1430 GCCATTGTGCTTTAATGACTGTATTGGGTATTTCATTAAACAATATCTTCATCATTTT 1489
Db 186 ValLeuAlaValLeuAlaValValGlyTyrSerLeuAsnAspThrIleValIlePhe 205
QY 1490 GATCGTATTCGTGAAGAT-----CGCCAAGCGAACCTGTTTACCCCTATGAT 1537
Db 206 AspArgValArgGluAsnPheArgValLeuArgLysAlaAspLeuValGluAsnLeuAsn 225
QY 1538 GTTTTAGTTAATGATGCCCTTCAAAAGAGCTTCAGCCGACGGTAAGTACAAACAGCTACA 1597
Db 226 Ile-----SerThrSerGlnThrLeuLeuArgThrIleAlaThrSerValSer 241
QY 1598 ACTCTATCAGTTTGTGTTAATGCTTTTATAGCGGCTCTCTGCTCTTTAATTTGCA 1657
Db 242 ThrLeuLeuAlaIleAlaAlaLeuLeuPhePheGlyGlyAspAsnLeuPheGlyPheSer 261

QY 1658 TTATTATGACCATAGGATTCTCTAGGAACCTTATCGTCTCTTATATATGACCACT 1717
Db 262 IleAlaLeuPheValGlyValMetAlaGlyThrTyrSerSerIleTyrIleAlaAsnVal 281
QY 1718 CTGTTGTTGTTTATG 1732
Db 282 ValLeuIleTrpLeu 286

RESULT 34
Q9KLOQ PRELIMINARY; PRT; 612 AA.
AC Q9KLOQ;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Protein-export membrane protein SecD.
GN VCA0693.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Emdolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ustebach T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483 (2000).
DR EMBL; AB004398; AAF96592.1; -.
DR TIGR; VCA0693;
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF; 1.
DR TIGRFAMs; TIGR00916; 2A0604801; 1.
DR TIGRFAMs; TIGR01129; secD; 1.
KW Complete proteome.
SQ SEQUENCE 612 AA; 66894 MW; B0466FF76C175E8E CRC64;

Alignment Scores:
Pred. No.: 1,1e-22 Length: 612
Score: 368.00 Matches: 79
Percent Similarity: 59.17% Conservative: 63
Best Local Similarity: 32.92% Mismatches: 94
Query Match: 11.11% Indels: 4
DB: 16 Gaps: 3

US-09-868-987-1 (1-1864) x Q9KLOQ (1-612)
QY 38 ACTGCTAATGACATATCTGCAACCGTGGATGGCTATGCTGTAGTGTGACGGT 97
Db 370 ThrValTyrArgGluTyrLysThrAsnAla-----ArgGlyGluThrValArgSerGlu 387
QY 98 TATATGCTGACGAGCCCTATTTTAAACGTCCTTAAACATCATGCCAGTGTCTCAGGG 157
Db 388 LysValIleSerValIleThrIleGlnSerGlnLeuGlySerGlnPheArgIleThrGly 407
QY 158 AAATTTACCCACCGTGAAGTGAAGCAACTCGCTTCAGATTTTAAATCTGGAGCGATGCT 217
Db 408 AlaGlySerMetGluGluAlaGlnGlnLeuAlaLeuLeuLeuAlaGlySerLeuThr 427
QY 218 TTGTTCCCGAGGTTCTCAGTGAAGAGACGATCTTCTTCTGATCTTGGGAAAAACAATGT 277
Db 428 AlaProValThrIleValGluGluArgThrIleGlyAlaSerLeuGlyGluGluAsnIle 447
QY 278 ACACAGGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTATTGTTTTCATCAGCGTA 337
Db 448 AlaAsnGlyPheAlaAlaLeuAlaLeuGlyMetAlaMetThrLeuThrPheMetAlaLeu 467
```

```

QY 338 TATATAGATTGGAGGCGCTATGCTCGGAGGCTGTTCTTGAATCTTTGCTTATC 397
Db 468 TPTPTATGATGAGTGTATPVALAALASNVALLALEUVALASNMETALCYSEU 487
QY 398 TGGGACACTCTACAGTATTTG---GATGCGCACTCACTTGTGAGACTGCTGGGATT 454
Db 488 LeuGlYleuIleAlaLeuLeuProGlyAlaValLeuThrLeuProGlyIleAlaGlyLeu 507
QY 455 GTTCTTGCTATGGAGTGGCGGTGATGCAATGTTCTTGATTGCAAGAAATCCGAG 514
Db 508 ValLeuThrValGlyMetAlaValAspThrAsnValLeuIlePheGluArgIleArgAsp 527
QY 515 GAATTTTATGTCTCAAGCTTAAACCTGAGAAAGATATACCAAGGCTTTT 574
Db 528 LysLeuSerGluGlyArgAsnPheAlaGlnAlaIleAspThrGlyPheAspSerAlaPhe 547
QY 575 GGAGCCATTTTGAATTCATCTTGACACTAGTATGGCTGAGACCTCTTTCTTCTTA 634
Db 548 SerThrIleLeuAspAlaAsnIleThrThrMetIleThrAlaValIleLeuTyrSerIle 567
QY 635 GATACAGGCGCTATTAAGGCTTGTGCTTGAACATGATTGAGAAATTTCTTCAATG 694
Db 568 GlyAsnGlyProIleGlnGlyPheAlaLeuThrLeuGlyLeuLeuThrSerMet 587
QY 695 TTTACGCTCTTTTCATGACTAAATTTTCTTC---ATGCTGTGATGAATTAAGCCCA 751
Db 588 PheSerGlyValPheAlaSerArgAlaLeuIleAsnLeuValTyrGlyArgAspThrArg 607

RESULT 35
Q9KDH7 PRELIMINARY; PRT; 437 AA.
AC Q9KDH7;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Protein-export membrane protein.
GN SECDF OR BH1236.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MBDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Maui N.,
RA Fuji F., Hixama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331 (2000).
DR EMBL; AP001511; BAB04955.1;
DR InterPro: IPR001036; AcRiflin res.
DR InterPro: IPR00335; SecD SecE.
DR Pfam: PF02355; SecD; SecE; 1.
DR PRINTS; PR00702; ACRIFAVINRP.
DR TIGRFAMs; TIGR00916; 2A0604s01; 1.
DR TIGRFAMs; TIGR01129; secD; 1.
KM Complete proteome.
SQ SEQUENCE 437 AA; 47822 MW; 5799F8994787DB38 CRC64;

Alignment Scores:
Pred. No.: 1.73e-22 Length: 437
Score: 365.50 Matches: 80
Percent Similarity: 56.92% Conservative: 68
Best Local Similarity: 30.77% Mismatches: 99
Query Match: 11.04% Indels: 13
DB: 16 Gaps: 4

US-09-868-987-1 (1-1864) x Q9KDH7 (1-437)
QY 20 CAGAGGGGATCAGCGGACGCTPAAATGACACATTTCTGCAACCGTGATGGGTATG 79
:::|||||:::
:::|||||:::

```

```

Db 182 GluGluGlyValAlaSerTyrAlaGluAlaMetLysAlaSerProLysTyr----- 199
QY 80 GCTGTAGTATTGACGGTTATATGTCAGACGCCATTTTA-----AACGCCATTG 133
Db 200 -----IleSerAlaAlaSerValAsnGlnProLeuPheThrArgAspValMetIle 216
QY 134 AAAATCATCCAGTGTCTCAAGGAAATTTACCCAGCGTGAAGTGAACAACTCGGCTCA 193
Db 217 GluAsnArg-----SerGlyAspPheThrValGluGlnThrArgPheIleAlaAsp 233
QY 194 GAATTAATCTGAGAGGATGCTTTGTTCCGAGGTTCTCAGTGAAGAGCATCTCT 253
Db 234 IleLeuAsnAlaGlySerLeuProAlaLysLeuAspValLeuSerSerAsnSerValSer 253
QY 254 TCGATCTTGGGAAAAAACAATGACACAGCATTAATCTCAGCAATGCTGTGGCGCA 313
Db 254 AlaSerLeuGlyGluGlnAlaMetGluArgThrValTyrAlaGlyPheIleGlyValAla 273
QY 314 ATGCTTATTGTTTGAATGACGATATTAATGATTTGAGAGGCGCTACGCTCGGAGCT 373
Db 274 LeuIlePheLeuTyrMetLeuLeuTyrTyrArgPheMetCylMetIleAlaThrIleThr 293
QY 374 GTTCTTGATGATCTTGTGCTTATCTGGGAGCTCTACAGTATTGGAGCCACTCAC 433
Db 294 LeuThrValTyrIleTyrLeuValLeuLeuIlePheAsnThrMetAsnAlaValLeuThr 313
QY 434 TTGTCAAGACTCGCTGGGATGTTGTTCTGATGGGAGTGGCGGTAGTCAAAATGTTCTT 493
Db 314 LeuProGlyIleAlaIleAlaLeuIleLeuGlyValGlyMetAlaValAspAlaAsnIleIle 333
QY 494 GTATGCAAGAAATCCGAGAGCAATTTTATGTCCTCAAGTCTTAAAGATGTAGAA 553
Db 334 ThrTyrGluArgIleLysGluGlnIleArgSerGlyLysSerIleMetSerSerPheLys 353
QY 554 AAGGATATACCAAGGCTTTGGAGCCATTTGATTCATCTGACCTACAGTATGGCC 613
Db 354 AlaGlySerArgArgSerLeuSerThrIleLeuAspAlaAsnIleThrThrIleLeuAla 373
QY 614 TCAGCACTTTTCTTCTCTGATGACAGGCGCTATTAAAGGTTTGCTTGACATTGATT 673
Db 374 AlaSerValLeuTyrIleTyrGlyThrSerSerValGlnIlyPheAlaValMetLeuIle 393
QY 674 TTAGGAATTTCTCTCAATGTTTACGCTTCTTCAATGATCAATATTTTCTTCATGCTG 733
Db 394 ValSerIleLeuMetSerPheValThrAlaValPheIleSerArgLeuLeuGlyLeu 413
QY 734 TGGATGATTAAGACCAACATACACAGTTGATGATGATGATGATGATGATGATGATG 793
Db 414 TrpValAsnSer-----ArgThrLeuAsnLysLeuTyrArgLeuPheGlyValLys 430

RESULT 36
Q8XGX9 PRELIMINARY; PRT; 323 AA.
AC Q8XGX9;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Peptide protein translocase, IISp family, membrane subunit (Protein-export
DE membrane protein SecE).
GN SECDF OR STM0408 OR STM0446.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SCS0412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Lattelle P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,

```

RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 L72.";  
 RL Nature 413:852-856(2001).  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RX SPECIES=S.typhi; STRAIN=CT18;  
 MDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,  
 RA Baker S., Basham D., Brooks K.K., Chillingworth T., Connerton P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,  
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrall B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18.";  
 RL Nature 413:848-852(2001).  
 DR EMBL; AE008714; AAU19362.1; -;  
 DR EMBL; AL627266; CAD08864.1; -;  
 DR InterPro; IPR003335; SecD\_SecF.  
 DR Pfam; PF02355; SecD\_SecF; 1.  
 DR TIGRFAMs; TIGR009167; 2A0604s01; 1.  
 DR TIGRFAMs; TIGR00966; 3A0501s07; 1.  
 DR Complete proteome.  
 SQ SEQUENCE 323 AA; 35376 MW; 10D06394DF60CB86 CRC64;

Alignment Scores:  
 Pred. No.: 1.85e-22 Length: 323  
 Score: 365.00 Matches: 103  
 Percent Similarity: 49.68% Conservative: 53  
 Best Local Similarity: 32.80% Mismatches: 122  
 Query Match: 11.02% Indels: 36  
 DB: 16 Gaps: 11

US-09-868-987-1 (1-1864) x Q8XGX9 (1-323)

Qy 794 CATGATTCTTTGAGAGATGCAAAAACCTTTGGGCT-----GTTCTCGAAGTGTTTT 847  
 Db 17 TrpAepPheMetArgTrp-----AspPheTrpAlaPheGlyIleSerGlyLeuLeuLeu 34  
 Qy 848 CTTTATAGTTCGGTCTGCTCGGTTTGGAGCCTGGAATTCGGTTTGGGAATGATTTT 907  
 Db 35 IleAlaAlaIleValIleMetGlyValArgGlyPheAsn-----TrpGlyLeuAspPhe 52  
 Qy 908 AAGGAGGATGCTCTTTACCTTTAATCAAAAGAGATGCGATCAGCGATGCTCTCAA 967  
 Db 53 ThrGlyGlyThrValIleGluIleThrLeuGluIysProAlaGluMetAspVal----- 70  
 Qy 968 ATGGGTGCCAAAGTTGTGCATAAACTACAGAAAGCTGGTCTTCTCTAGAGACTTCGT 1027  
 Db 71 MetArg-----GluAlaLeuGlnIysAlaGlyTyr-----GluGluProGln 84  
 Qy 1028 ATTCAACATTGGATCTTCAGAAAGATCAAAATCTATTATTAGTGATAAGCTTTAAGC 1087  
 Db 85 LeuGlnAsnPheGlySerSerHisAspIleMetValArgMetProThrGluGlyGlu 104  
 Qy 1088 TATACTAGCAGATACAGCTCTCTCTCTAAATTAAGCATCATGAGCTGGCGTTATTGT 1147  
 Db 105 ThrGlyGlyGlnValLeuGlySerIysVal----- 114  
 Qy 1148 GGGATTGTGTGCAGAAACAGGCTAGATTTCTCTACGGAACTCTAAACGAAACGCAAAA 1207  
 Db 115 ---ValThrIleIleAsnGluAlaThr-----AsnGlnAsnAlaValIys 129  
 Qy 1208 TTTTGGTCAAAGGTAAAGCAGCAAACTATCAGAAATAATCGGTATACGGCAGCATCGGG 1267  
 Db 130 ArgIleGluPheValGlyProSerValGlyAlaAspLeuAlaGlnThrGlyAlaMetAla 149  
 Qy 1268 CTTTATAGAGCTTTGGCAATCATCTGCTCTCTATGAGCTTTGGCTTTGAATGCAATAT 1327  
 Db 150 LeuLeuValAlaIleSerIleLeuValTyrValGlyPheArgPheGluTrpArgLeu 169

Qy 1328 GCTTTCAGTCCGCTATGCGCTTTAATTAATCATGACCTTTTGGCTACCTGCGAGTCTTGT 1387  
 Db 170 AlaAlaGlyValValIleAlaLeuAlaHisAspValIleThrLeuGlyIleLeuSer 189  
 Qy 1388 ATAGCACATTTCTTTTGAAGAAAATCAATAGATTTGCAAGCCATTGGTGCCTTTAATG 1447  
 Db 190 LeuPheHis-----IleGluIleAspLeuThrIleValAlaSerLeuMet 204  
 Qy 1448 ACTTATTTGGGTATTTCATTAACAATACCTTTGATCATCTTTTGTATTCGTGGAAGAT 1507  
 Db 205 SerValIleGlyTyrSerLeuAsnAspSerIleValValSerAspArgIleArgGluAsn 224  
 Qy 1508 ---CGCCAAGCAACCTGTTTACCCTATGATGATTTTATGATGATGATGATGATGATG 1564  
 Db 225 PheArgIysIleArgArgGlyThrProTyrGluIlePhe---AsnValSerLeuThrGln 243  
 Qy 1565 ACCTTCAGCGCAGGTAATGACACCACTACACTATCATCATCTTTGTTGTAATGCTTTG 1624  
 Db 244 ThrLeuHisArgThrLeuIleThrSerGlyThrThrLeuValValIleLeuMetLeuTyr 263  
 Qy 1625 TTTATAGCGGCTCTCTGCTCTTTAATTTTGCATTTATTATGACCATAGGATCTTCTA 1684  
 Db 264 LeuPheGlyGlyProValLeuGluGlyPheSerLeuThrMetLeuIleGlyValSerIle 283  
 Qy 1685 GGAACTTTATGCTCTCTTTATTTGACCACTACCTCTGTTGTTG 1726  
 Db 284 GlyThrAlaSerSerIleTyrValAlaSerAlaLeuAlaLeu 297

RESULT 37

Q9K0J0 ID Q9K0J0 PRELIMINARY; PRT; 618 AA.  
 AC Q9K0J0;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Protein-export membrane protein SecD.  
 GN NMB0607.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=491;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / SEROGROUP B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.P., Dodson R.J.,  
 RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,  
 RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,  
 RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 RT MC58.";  
 RL Science 287:1809-1815(2000).  
 DR EMBL; AE002416; AAP41034.1; -;  
 DR TIGR; NMB0607; -;  
 DR InterPro; IPR003335; SecD\_SecF.  
 DR Pfam; PF02355; SecD\_SecF; 1.  
 DR TIGRFAMs; TIGR00916; 2A0604s01; 1.  
 DR TIGRFAMs; TIGR01129; secD; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 618 AA; 66847 MW; F4F805FB28B0AF10 CRC64;

Alignment Scores:  
 Pred. No.: 2.4e-22 Length: 618  
 Score: 364.00 Matches: 80  
 Percent Similarity: 55.95% Conservative: 61  
 Best Local Similarity: 31.75% Mismatches: 105  
 Query Match: 10.99% Indels: 6  
 DB: 16 Gaps: 3

US-09-868-987-1 (1-1864) x Q9JUV5 (1-618)

QY 23 GAGGGGATCAGCGGCACTGCTAATGACAAATAT--TCTGCAAAACCGTGATGGCGTATG 79  
 Db 348 AsperglaaglyserilepnehlleuThralaalaenValgilylaargmet 367  
 QY 80 GCTGTACTGAT-----GAGCGTATATGCGACAGCCGCTATTAAAGCTC 127  
 Db 368 AlalethralleuileaspGingilylserleuValValIntrAlaProvalIleargThr 387  
 QY 128 CCATTGAAA--AATCATGCCAGTGTCTCAGGAAATTTACCAACCGTAAGTACGAAA 184  
 Db 388 AlalethralleuileaspGingilylserleuValValIntrAlaProvalIleargThr 407  
 QY 185 CTCGCCATGATTTAAATCTGAGCGATGCTTTGTTCCGAGGTTCTCAGTGAAGAG 244  
 Db 408 ThrieuThrieuLeuArgAlaGlySerleuAlaIaPrometGlnIleValIgluIarg 427  
 QY 245 ACGATCTCTTCTGATCTTGGGAAAAACATGTACACAGGCAATTATCTCAGCATGCTGT 304  
 Db 428 ThrieuThrieuLeuArgAlaGlySerleuValValIntrAlaProvalIleargThr 447  
 QY 305 GCGTTGGCAATGCTTATGTTGATGAGCGTATATTTAGATTGAGGCGTATCGCT 364  
 Db 448 ThrieuThrieuLeuArgAlaGlySerleuValValIntrAlaProvalIleargThr 467  
 QY 365 TCGGAGCGTGTCTTGAATCTTTGCTTATCTGCGGAGCTCTACAGTATTTGGATGCG 424  
 Db 468 ThrieuThrieuLeuArgAlaGlySerleuValValIntrAlaProvalIleargThr 487  
 QY 425 CCACTGACCTGTGATGAGCGTGTGCTGCTGCTGATGAGGCGGCTGATGATGCA 484  
 Db 488 ThrieuThrieuLeuArgAlaGlySerleuValValIntrAlaProvalIleargThr 507  
 QY 485 AATGTTCTTGTATGCAAGATCCGAGAGAAATTTATTTGCTCAAGTCTTAAAAA 544  
 Db 508 AsnValleuileaspGingilylserleuValValIntrAlaProvalIleargThr 527  
 QY 545 TCTGTAGAAAAAGATATACCAAGGCTTTTGAGGCAATTTTATTTGATTTCACTGATCA 604  
 Db 528 AlalethralleuileaspGingilylserleuValValIntrAlaProvalIleargThr 547  
 QY 605 GTATTGGCGTACAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 664  
 Db 548 LeuileaspGingilylserleuValValIntrAlaProvalIleargThr 567  
 QY 665 ACATTGATTTAGGAAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 724  
 Db 568 ValhiscysleuGingilylserleuValValIntrAlaProvalIleargThr 587  
 QY 725 TTCACTGCTGTGATGATTAAGCCCAACATACACAG 760  
 Db 588 ValhiscysleuGingilylserleuValValIntrAlaProvalIleargThr 599

RESULT 38

Q9JUV5 PREDIMINARY; PRT; 618 AA.

AC Q9JUV5; TREMBL; 15, Created

DT 01-OCT-2000 (TREMBL; 15, Last sequence update)

DT 01-JUN-2002 (TREMBL; 21, Last annotation update)

DE Hypothetical protein SEC2 OR NMA0812.

GN NMA0812.

OS Neisseria meningitidis (serogroup A).

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI\_TaxId=65699;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;

RX MEDLINE=20222556; PubMed=10761919;

RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,

RA Jagers K., Leach S., Moulé S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., RA Whitehead S., Spratt B.G., Barrall B.G.; "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491." RT Nature 404:502-506 (2000).

DR EMBL: AL162754; CAB84094.1; -. DR InterPro: IPR003335; SecD\_SecE. DR Pfam: PF02355; SecD\_SecE. DR TIGRPFAM: TIGR00916; 2A0604601.1. DR TIGRPFAM: TIGR01129; secD; 1. DR Hypothetical protein; Complete proteome. SQ SEQUENCE 618 AA; 66888 MW; 138AE6760A79BD8 CRC64;

Alignment Scores:

Pred. No.	Length:	Score:	Matches:	Conservative:	Mismatches:	Query Match:	Indels:	Gaps:
2,4e-22	618	364.00	80	55.95%	31.75%	10.99%	6	3

US-09-868-987-1 (1-1864) x Q9JUV5 (1-618)

QY 23 GAGGGGATCAGCGGCACTGCTAATGACAAATAT--TCTGCAAAACCGTGATGGCGTATG 79  
 Db 348 AsperglaaglyserilepnehlleuThralaalaenValgilylaargmet 367  
 QY 80 GCTGTACTGAT-----GAGCGTATATGCGACAGCCGCTATTAAAGCTC 127  
 Db 368 AlalethralleuileaspGingilylserleuValValIntrAlaProvalIleargThr 387  
 QY 128 CCATTGAAA--AATCATGCCAGTGTCTCAGGAAATTTACCAACCGTAAGTACGAAA 184  
 Db 388 AlalethralleuileaspGingilylserleuValValIntrAlaProvalIleargThr 407  
 QY 185 CTCGCCATGATTTAAATCTGAGCGATGCTTTGTTCCGAGGTTCTCAGTGAAGAG 244  
 Db 408 ThrieuThrieuLeuArgAlaGlySerleuAlaIaPrometGlnIleValIgluIarg 427  
 QY 245 ACGATCTCTTCTGATCTTGGGAAAAACATGTACACAGGCAATTATCTCAGCATGCTGT 304  
 Db 428 ThrieuThrieuLeuArgAlaGlySerleuValValIntrAlaProvalIleargThr 447  
 QY 305 GCGTTGGCAATGCTTATGTTGATGAGCGTATTTATTTGATTTGAGGCGTATCGCT 364  
 Db 448 ThrieuThrieuLeuArgAlaGlySerleuValValIntrAlaProvalIleargThr 567  
 QY 365 TCGGAGCGTGTCTTGAATCTTTGCTTATCTGCGGAGCTCTACAGTATTTGGATGCG 424  
 Db 468 ThrieuThrieuLeuArgAlaGlySerleuValValIntrAlaProvalIleargThr 487  
 QY 425 CCACTGACCTGTGATGAGCGTGTGCTGCTGCTGATGAGGCGGCTGATGATGCA 484  
 Db 488 ThrieuThrieuLeuArgAlaGlySerleuValValIntrAlaProvalIleargThr 507  
 QY 485 AATGTTCTTGTATGCAAGATCCGAGAGAAATTTATTTGCTCAAGTCTTAAAAA 544  
 Db 508 AsnValleuileaspGingilylserleuValValIntrAlaProvalIleargThr 527  
 QY 545 TCTGTAGAAAAAGATATACCAAGGCTTTTGAGGCAATTTTATTTGATTTCACTGATCA 604  
 Db 528 AlalethralleuileaspGingilylserleuValValIntrAlaProvalIleargThr 547  
 QY 605 GTATTGGCGTACAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 664  
 Db 548 LeuileaspGingilylserleuValValIntrAlaProvalIleargThr 567  
 QY 665 ACATTGATTTAGGAAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 724  
 Db 568 ValhiscysleuGingilylserleuValValIntrAlaProvalIleargThr 587  
 QY 725 TTCACTGCTGTGATGATTAAGCCCAACATACACAG 760

```

Db 588 ValAsnLeuTrpTyrglyArgArgLysLeuGln 599
RESULT 39
O9KTY6
ID O9KTY6 PRELIMINARY; PRT; 315 AA.
AC O9KTY6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-export membrane protein SecF.
GN VC0744.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hick E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ueberback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AB004160; AAF93909.1; -.
DR TIGR; VC0744; -.
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF; 1.
DR TIGRFAMs; TIGR00916; 2A0604s01; 1.
DR TIGRFAMs; TIGR00966; 3A0501s07; 1.
KW Complete proteome.
SQ SEQUENCE 315 AA; 34329 MW; 634AF1AFFC070BD CRC64;

Alignment Scores:
Pred. No.: 2.48e-22 Length: 315
Score: 363.50 Matches: 104
Percent Similarity: 50.62% Conservative: 59
Best Local Similarity: 32.30% Mismatches: 104
Query Match: 10.98% Indels: 55
DB: 16 Gaps: 13

US-09-868-987-1 (1-1864) x O9KTY6 (1-315)
Qy 797 GATTTCCTGAGAGGATGCAAAACCTTTGGGCTGTTCT-----GGAAAGT 841
Db 12 AspPheMetArgTrpSerLysPheAlaPheAlaLeuSerLeuValMetIleAlaAlaSer 31
Qy 842 GTTTTCCTTTAGGTGCTGCTCGGTTTGAGGCTGGAATTCGTTTGGGAATG 901
Db 32 IlePheThrLeuSerThrLysTrpLeuAsn-----Trp-----GlyLeu 44
Qy 902 GATTTTAAAGAGGGTATGCGCTTTACCTTTTAATCCAAAGAGCATGGCATGCT 961
Db 45 AspPheThrGlyGlyThrLeuIle-----GluValGlyPheGluGlnPro 59
Qy 962 GCTCAAAATGCGTGGCAAGTGTGTGATCAAACTACAGGAGCTGGTCTTCTTCTAGAC 1021
Db 60 AlaAsnLeuGlu---GlnIleArgSerAlaLeuGluAlaLysGlyPheGly-----Asp 76
Qy 1022 TTCCTATTCAACATTTGGATCTTCAGAAAGATCAAAATCTAT----- 1066
Db 77 AlaThrValGlnAsnPheGlySerAlaArgAspValMetValArgLeuArgProArgAsp 96
Qy 1067 ---TTTAGTATAAAGCTTTAAGCTATATAAGACAGATACGAGCTCTCTCTCTAAATA 1123
Db 97 AspMetAlaGlyGluAlaLeuGly-----AsnGlnIleLeuAlaAlaIleLysGly 114
Qy 1124 ACGATCATGAGCTGGCGTTATTGTGGGATTGTGTGAGAAACAGCGCTAGATTCTCTAC 1183

```

```

Db 115 Thr-----GlyGlyAsnValGluMetArgArgIleGluPhe----- 126
Qy 1184 GGAAGCTCTAAACGAAACGAAATTTTGGTCAAGGTAAGACGACAACTATCGAAGAAA 1243
Db 127 -----ValGlyProAsnValGlyAspGlu 134
Qy 1244 ATGCGTTATCAGCGACCATCGGGCTTTTAGGAGCTTTTGGCAATCATCTTGCTCTATGTG 1303
Db 135 LeuThrGluAlaGlyGlyLeuAlaIleLeuValSerLeuLeuCysIleLeuLeuVal 154
Qy 1304 AGTTTGGCTTTTGAATGCAATATGCTTTTTCAGTCCGCTATGCGCTTTAATTCATGACCTT 1363
Db 155 SerValArgPheGluTrpArgLeuAlaAlaGlyAlaValLeuAlaLeuAlaHisAspVal 174
Qy 1364 TTGGCTACCTGTGCAGCTGTTTATATGACCATTTCTTTTGAAGAAAATTCAAATAGAT 1423
Db 175 IleIleThrLeuGlyIle-----PheSerIleLeuGlnIleGluValAsp 189
Qy 1424 TTGCAAGCCATTGGTCTTTAATGACTGTATGGGGTATTTCATTAAACAATCTTTGATC 1483
Db 190 LeuThrIleValAlaAlaLeuLeuThrValValGlyTyrSerLeuAsnAspThrIleVal 209
Qy 1484 ATTTTTCATCGTATTCTCGTAGAT---CGCCAAGCGAACCTGTTTACCCCTATCATGTT 1540
Db 210 ValPheAspArgIleArgGluAsnPheArgLysGluGluProAlaGluIle 229
Qy 1541 TTAGTTAATGATGCCCTTCAAAAGACGCTTCACGCGACGTAATGACAAACAGCTACAACT 1600
Db 230 Met---AsnSerSerIleThrGlnThrLeuSerArgThrLeuIleThrSerGlyThrThr 248
Qy 1601 CTATCAGCTTTTGTAAATGCTTTTGTATATAGGCGGCTCTCTGCTTTAATTTTGCATTT 1660
Db 249 LeuPheValValIleAlaLeuPheThrGlnGlyGlyAlaMetIleHisGlyPheAlaLeu 268
Qy 1661 ATATGACCATAGGATCTTCTAGGAACCTTATCGTCTCTTATATATGACCACTCTG 1720
Db 269 AlaLeuLeuLeuGlyIleThrValGlyThrTyrSerSerIleTyrValAlaSerAlaLeu 288
Qy 1721 TTGTGT 1726
Db 289 AlaLeu 290
RESULT 40
O9PGS3
ID O9PGS3 PRELIMINARY; PRT; 614 AA.
AC O9PGS3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-export membrane protein.
GN XF0225.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,

```

RA Marques M.V., Martins E.F.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.B.S.,  
 RA Nham A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Zeldman J., Zerbini J.C.,  
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"  
 RL Nature 406:151-159(2000).  
 DR EMBL: AEO03876; AAF83038.1; -  
 DR InterPro: IPR001036; Acrlivin.res.  
 DR InterPro: IPR003335; Secd SecF.  
 DR Pfam: PF02355; Secd SecF\_1.  
 DR PRINTS: PR00702; ACRIPLAVNRP.  
 DR TIGRPFAMS: TIGR00916; 2A0604501; 1.  
 DR TIGRPFAMS: TIGR01129; secf, 1.  
 KM Complete proteome.  
 SQ SEQUENCE 614 AA; 6699 MW; C8EB17ASEDONAFAS CRC64;

Alignment Scores:  
 Pred. No.: 8.53e-22 Length: 614  
 Score: 357.50 Matches: 80  
 Percent Similarity: 55.36% Conservative: 44  
 Best Local Similarity: 35.71% Mismatches: 83  
 Query Match: 10.79% Indels: 17  
 DB: 16 Gaps: 3

US-09-868-987-1 (1-1864) x Q9PGS3 (1-614)

QY 86 GTGATGCGGTATATGTCAGCAGCCCTATTTAAACATTCATGGAATAATCAT--- 142  
 Db :|||||  
 377 MetValaspGly-----LysGluValArgGlyValArgLeuValGluGluVal 392  
 QY 143 -----GTCAGTGTCTCAGGAAA-----TTTACC 166  
 Db :|||||  
 393 LeuSerProThrArgGlyLeuValArgGlyValArgLeuValGluGluVal 412  
 QY 167 CACCGTGAAGTGAACAACCTGCTCAGATTAAATCGAGGAGATCTTTTGTCCC 226  
 Db :|||||  
 413 LysThrGluValGluValSerLysLeuValArgGlySerLeuValAlaProMet 432  
 QY 227 GAGGTTCGATGGAAGAAGACATCTCTGATCTTGGAAGAAAACATGATACAGAGC 286  
 Db :|||||  
 433 AspHeuValGluGluValrValIleGlyProSerLeuGlyAlaGluValArgGly 452  
 QY 287 ATTATCTCAGCATGCTGTGCTTGGCAATGCTTATTTGATGAGGATATATAGA 346  
 Db :|||||  
 453 ValThrAlaValAlaPheSerPheLeuPheThrLeuValPhePheThrIleTyrTyrArg 472  
 QY 347 TTTCGAGCGTCATCGCTTGGAGCTCTTCTGCAATCTTTTCTTATCTGGACACT 406  
 Db :|||||  
 473 MetPheGlyValIleThrSerLeuAlaLeuLeuPheAsnLeuIleValAlaVal 492  
 QY 407 CTACAGATTGGAATGGCAGCTCAGCTTGAGGACTCGCTGGATGTTCTTGCTATG 466  
 Db :|||||  
 493 MetSerLeuPheGlyAlaThrMetThrLeuProGlyPheAlaGlyLeuAlaLeuSerVal 512  
 QY 467 GGGATGGCCGTAGATGCAATGTTCTTGTATTCGAAAGATCCGAGAGAAATTTTATG 526  
 Db :|||||  
 513 GlyLeuSerValAlaPheAlaAsnValLeuIleAsnGluArgIleArgGluGluLeuArgHis 532  
 QY 527 TCTCAAGTCTTAAATACTGTAGAAAAGATATACCAAGCTTTTGGACCATTTT 586  
 Db :|||||  
 533 GlyMetProPheArgSerAlaIleValAlaGlyTyrAspValAlaGlySerThrIleLeu 552  
 QY 587 GATCTTAACTGATCTACAGTATGCGCTCAGACATCTTTCTTCTTATGATACAGGCGCT 646  
 Db :|||||

Db 553 AspAlaAsnLeuThrGlyLeuIleValAlaAlaLeuTyrAlaPheGlyThrGlyPro 572  
 QY 647 ATTTAAGCGTTTCTTGATGATTTTGAAGATTTTCTTTCATGTTTACGCTCTT 706  
 Db :|||||  
 573 LeuValGlyPheAlaLeuThrMetIleIleGlyIlePheAlaSerMetPheThrAlaIle 592  
 QY 707 TTCATGACTTAA 718  
 Db :|||||  
 593 ThrValSerArg 596

RESULT 41  
 ID Q9RNL9 PRELIMINARY; PRT; 615 AA.  
 AC Q9RNL9;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE General protein secretion pathway subunit SecD.  
 OS Citrobacter freundii.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Citrobacter.  
 OX NCBI\_TaxID=546;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kuhn F.C., Sobieski R.J., Crupper S.S.;  
 RT "Citrobacter freundii secD gene,"  
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF179925; AAD53945.1; -  
 DR InterPro: IPR003335; Secd SecF.  
 DR InterPro: IPR005058; SigTase.  
 DR Pfam: PF02355; Secd SecF\_1.  
 DR TIGRPFAMS: TIGR00916; 2A0604501; 1.  
 DR TIGRPFAMS: TIGR01129; secf, 1.  
 DR PROSITE: PS00761; SPASE\_13; UNKNOWN\_1.  
 DR PROSITE: PS00761; SPASE\_13; UNKNOWN\_1.  
 SQ SEQUENCE 615 AA; 66572 MW; DD96134077301547 CRC64;

Alignment Scores:  
 Pred. No.: 8.53e-22 Length: 615  
 Score: 357.50 Matches: 81  
 Percent Similarity: 61.50% Conservative: 58  
 Best Local Similarity: 35.84% Mismatches: 84  
 Query Match: 10.79% Indels: 3  
 DB: 2 Gaps: 2

US-09-868-987-1 (1-1864) x Q9RNL9 (1-615)

QY 44 AATGACAAATATTCGCAACCGTGATGCGCTGATGATGATGATGATGATGATG 103  
 Db :|||||  
 375 SerGlyLysLysAspAlaAsn-----GlyArgAlaValLeuValLysGluGluVal 392  
 QY 104 GTACAGAGCCCTATTTTAAACGTCATTCGAAATAATCATGCCGTCTCAGGAAATTT 163  
 Db :|||||  
 393 IleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleAsn 412  
 QY 164 ACCCAGCGTGAAGTGAACAACCTGCTCAGATTAAATCTGAGAGGATGCTTTTGT 223  
 Db :|||||  
 413 AsnProAsnGluAlaArgGluMetSerLeuLeuValArgIleValAlaLeuIleAlaPro 432  
 QY 224 CCGAGGTTCTCAGTGAAGAAGACATCTTCTGATCTTGGAAGAAAACATGATACAA 283  
 Db :|||||  
 433 IleGlnIleValGluGluValArgThrIleGlyProThrLeuGlyMetGlnAsnIleLysGln 452  
 QY 284 GGCATTTCTCAGACATGCTGTGGCTTGGCAATCTTATTTTGTGATGAGCGATATAT 343  
 Db :|||||  
 453 GlyLeuGluValAlaCysLeuAlaGlyLeuValAlaSerIleLeuPheMetIlePhePheTyr 472  
 QY 344 AGATTGAGCGGTCATCGCTTGGAGGCTGCTTCTGATCTTGGAAGAAAACATGATACAA 403  
 Db :|||||  
 473 LysLysPheGlyLeuIleAlaThrSerAlaLeuIleAlaAsnLeuValLeuIleValGly 492  
 QY 404 GCTCTACGATTTTG---GATGGCCCATCTCAGCTTCTGAGACTCGCTGGATTTGCTT 460  
 Db :|||||  
 493 IleMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512

```

Qy 461 GCTATGGGATGGCCGTAGATCAAAATGTTCTTGTATTTCGAAAGAAATCCGAGAGGAATTT 520
Db 513 ThrLeuAlaValAlaAspAlaAsnValLeuIleAsnGluArgIleLysGluGluLeu 532
Qy 521 TTATTGTCTCAAGCTTAAAGAAATCTGTAGAAAAGGATATACCAAGCGCTTTTCGAGCC 580
Db 533 SerAsnGlyArgSerValGlnGlnAlaIleAsnGluGlyTyrAlaGlyAlaPheSer 552
Qy 581 ATTTTGTATTCTAATCTAGTACATAGTATTGGCCCTCAGCACTCTTTTCTCTCAGATACA 640
Db 553 IlePheAspAlaAsnIleThrThrLeuIleLysValIleIleLeuTyrAlaValGlyThr 572
Qy 641 GGGCTATTAAGGTTTTCCTTTCATGATGATTAGGAATTTTCTCTCAATGTTTACG 700
Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaThrSerMetPheThr 592
Qy 701 GCTCTTTTTCATGACTAAA 718
Db 593 AlaIleValGlyThrArg 598

RESULT 42
Q8ZRD7 PRELIMINARY; PRT; 615 AA.
AC Q8ZRD7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Preprotein translocase, IISF family, part of the channel.
GN SEC OR STM0407
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=L72 / SGSC1412 / ATCC 700720;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT L72."
RL Nature 413:852-856(2001).
DR EMBL; AE008714; AAL19361.1; -.
DR InterPro; IPR003335; SecD_SecP.
DR InterPro; IPR000508; SigPase.
DR Pfam; PF02335; SecD_SecP; 1.
DR TIGRFAMs; TIGR00916; 2A0604s01; 1.
DR TIGRFAMs; TIGR01129; secD; 1.
DR PROSITE; PS00761; SPASE_1_3; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 615 AA; 66587 MW; 38383148F3916B33 CRC64;

Alignment Scores:
Pred. No.: 1.04e-21 Length: 615
Score: 356.50 Matches: 83
Percent Similarity: 59.75% Conservative: 61
Best Local Similarity: 34.44% Mismatches: 94
Query Match: 10.76% Indels: 3
DB: 16 Gaps: 2

US-09-868-987-1 (1-1864) x Q8ZRD7 (1-615)
Qy 44 AATGACAAATATCTGCAACCGTGGATGGCGTATGCGTATGACGATGACGATATATG 103
Db 375 SerGlyLysLysAspAlaAsn-----GlyArgAlaValLeuValLysGlnGluVal 392
Qy 104 GTCAGACGCTATTTTAAACGTCCTCCATGGAATATCATGCGATGCTCAGGGAATTT 163
Db 393 IleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleSer 412

```

```

Qy 164 ACCCACCGTGAAGTGAAGCAAACTCGCCTCAGATTTAAATCTGGAGCGATGCTTTTGT 223
Db 413 AsnProAsnGluAlaArgGlnSerLeuLeuLeuArgAlaGlyAlaLeuIleAlaPro 432
Qy 224 CCCAGGTTCTCAGTGAAGAGACCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 283
Db 433 IleGlnIleValGluGluArgThrIleGlyProThrLeuGlyMetGlnAsnIleLysGln 452
Qy 284 GGCATTATCTCAGCATGCTGGGCTGGCAATGCTTATTGTTTGTATGAGCGTATATTAT 343
Db 453 GlyLeuGluAlaCysLeuAlaGlyLeuValSerIleLeuPheMetIlePhePheThr 472
Qy 344 AGATTGGAGCGTCATCGCTTCGGGAGCTGTTCTCTCTGAATCTTTTCTCTCTCTCTCT 403
Db 473 LysLysPheGlyLeuIleAlaThrSerAlaLeuValAlaAsnLeuValIleValGly 492
Qy 404 GCTCTACAGTATTG---GATGCGCCACTCACCTTGTCTGAGCACTCGCTGGGATGTTCT 460
Db 493 IleMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512
Qy 461 GCTATGGGATGGCCGTAGATGCAAAATGTTCTTCTTATTCGAAAGAAATCCGAGAGGA 520
Db 513 ThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGluArgIleLysGluGluLeu 532
Qy 521 TTATTGTCTCAAGCTTAAAGAAATCTGTAGAAAAGGATATACCAAGCGCTTTTGGAGCC 580
Db 533 SerAsnGlyArgThrValGlnGlnAlaIleAsnGluGlyTyrAlaGlyAlaPheSerSer 552
Qy 581 ATTTTGTATCTCAACTTGACTACATGATGTCCTGAGCACTCTCTCTCTCTCTCTCTCTCT 640
Db 553 IlePheAspAlaAsnIleThrThrLeuIleLysValIleIleLeuTyrAlaValGlyThr 572
Qy 641 GGGCTATTAAGGTTTTCCTTTCACATGATTTTGTAGGAATTTTCTCTCTCAATGTTTACG 700
Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaThrSerMetPheThr 592
Qy 701 GCTCTTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAATAAGACCAACATACACAG 760
Db 593 AlaIleIleGlyThrArgAlaIleValAsnLeuLeuTyrGlyGlyLysArgValThrLys 612
Qy 761 TTG 763
Db 613 Leu 613

RESULT 43
Q8Z8X9 PRELIMINARY; PRT; 615 AA.
AC Q8Z8X9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Protein-export membrane protein secD.
GN STY0445.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CT18;
RY MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks R.M., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).

```



DR EMBL: AL627266; CND08863.1; -  
 DR InterPro: IPR003335; SecD\_SecF.  
 DR InterPro: IPR000508; SigPase.  
 DR Pfam: PF02355; SecD\_SecF.1.  
 DR TIGRPFAM: TIGR00916; 2A0604s01; 1.  
 DR TIGRPFAM: TIGR01129; secD.1.  
 DR PROSITE: PS00761; SPASE\_1.3; UNKNOWN\_1.  
 DR Complete proteome.  
 SW SEQUENCE 615 AA; 66605 MW; 81370B11FD5983F9 CRC64;

## Alignment Scores:

Pred. No.:	Length:	Matches:
Score: 356.50	615	83
Percent Similarity: 59.75%	Conservative: 61	
Best Local Similarity: 34.44%	Mismatches: 94	
Query Match: 10.76%	Indels: 3	
	Gaps: 2	

US-09-868-987-1 (1-1864) x Q828X9 (1-615)

QY 44 AATGACAAATATTCGCAACCGTGATGCGCTAGTGAATGACGTTATATG 103  
 Db 375 SerLysLysAspAlaAsn-----GlyArgAlaValLeuValLysGlnGluVal 392  
 QY 104 GTCAGCAGCCATTTTAAAGTCCCATGAAATCATGCGACATGCTCGGAAATTT 163  
 Db 393 IleAsnIleAlaAsnIleGlnSerArgLeuGlnSerPheArgIleThrGlyIleSer 412  
 QY 164 ACCACCGTGAAAGTGACAACTCGCCATGATTTAAATCTGAGCGATGCTTTGTT 223  
 Db 413 AsnProAsnGlnAlaArgGlnLeuSerLeuLeuArgAlaGlyAlaLeuIleAlaPro 432  
 QY 224 CCCAGGTTCTCAGTGAAGAGACATCTCTTGATCTGGGAAAAACAATGATACAA 283  
 Db 433 IleGlnIleValGlnGluArgThrIleGlyProThrLeuGlnMetGlnAsnIleLysGln 452  
 QY 284 GGCATTATCTCAGATGCTGCTGGCATGCTTATGTTGATGAGGATATATAT 343  
 Db 453 GlyLeuGlnAlaCysLeuAlaGlyLeuValValSerIleLeuPheMetIlePhePheTyr 472  
 QY 344 AGATTGGAGGCGTCATCGGCTTGGAGAGCTTCTTGATCTTGATCTTGATCTG 403  
 Db 473 LysLysPheGlyLeuIleIleThrSerAlaLeuValAlaAsnLeuValIleValGly 492  
 QY 404 GCTCTACAGTATTTG--GATGCGCCACTCACTTTCAGACATCGCTGGATTTCTT 460  
 Db 493 IleMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512  
 QY 461 GCTATGGGATGCGCGTCATCGGCTTGGAGAGCTTCTTGATCTTGATCTTGATCTG 520  
 Db 513 ThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGlnArgIleLysGlnGluLeu 532  
 QY 521 TTATGTCGCAAGCTTAAATAATCTGTAGAAAAAGATATACCAAGCTTTGAGACC 580  
 Db 533 SerAsnGlyArgThrValClnGlnAlaIleAsnGlnGlyTyrAlaGlyAlaPheSerSer 552  
 QY 581 ATTTTGAATCTAATCTGATCTGATGATGAGCTTCTTCTTCTTCTTCTTCTTCTT 640  
 Db 553 IlePheAspAlaAsnIleThrThrLeuIleLysValIleIleLeuTyrAlaValGlyThr 572  
 QY 641 GGGCTTAAAGGTTGCTTGAATGATGATGATGATGATGATGATGATGATGATGATG 700  
 Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaThrSerMetPheThr 592  
 QY 701 GCTCTTTCATGAAATTTTCTTCTGATGATGATGATGATGATGATGATGATGATGATG 760  
 Db 593 AlaIleIleGlyThrArgAlaIleValAsnLeuLeuTyrGlyGlyLysArgValThrLys 612  
 QY 761 TTG 763  
 Db 613 Leu 613  
 RESULT 44

Q9XBN2 PRELIMINARY; PRT; 615 AA.

AC Q9XBN2  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE SecD protein.  
 GN SEC D  
 OS Enterobacter aerogenes (Enterobacter aerogenes).  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Enterobacter.  
 RX NCBI\_TaxID=548;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Kuhn F.C., Gies A.J., Smeltzer M., Crupper S.S., Sobieski R.J.;  
 RT "Identification of the secD gene of Enterobacter aerogenes."  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF163861; AAD44348.1; -  
 DR TIGRPFAM: TIGR00916; 2A0604s01; 1.  
 DR InterPro: IPR003335; SecD\_SecF.  
 DR Pfam: PF02355; SecD\_SecF.1.  
 DR TIGRPFAM: TIGR00916; 2A0604s01; 1.  
 DR PROSITE: PS00761; SPASE\_1.3; UNKNOWN\_1.  
 DR SEQUENCE 615 AA; 66744 MW; 6BA2C06542F715D3 CRC64;

## Alignment Scores:

Pred. No.:	Length:	Matches:
Score: 356.00	615	77
Percent Similarity: 63.77%	Conservative: 55	
Best Local Similarity: 37.20%	Mismatches: 69	
Query Match: 10.75%	Indels: 6	
	Gaps: 2	

US-09-868-987-1 (1-1864) x Q9XBN2 (1-615)

QY 116 ATTTTAAAGCTCCCA-----TTGAAAAATCATGCCAGTCTCAGGAGAA 160  
 Db 392 ValIleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIle 411  
 QY 161 TTACCCACCGTGAAAGTGACAACTCGCCATGATTTAAATCTGAGCGATGCTTTT 220  
 Db 412 SerAsnProThrGlnAlaArgGlnLeuSerLeuLeuArgAlaGlyAlaLeuIleAla 431  
 QY 221 GTTCCGAGGTTCTCAGTGAAGAGACATCTCTTGATCTTGATCTTGATCTG 280  
 Db 432 ProIleGlnIleValGlnGluArgThrIleGlyProThrLeuGlnMetGlnAsnIleLys 451  
 QY 281 CAAGGCTTATCTCAGATGCTGCTGGCATGCTTATGTTGATGAGCGATATAT 340  
 Db 452 GlnIleLeuGlnAlaCysLeuAlaGlyLeuValValSerIleLeuPheMetIlePhePhe 471  
 QY 341 TATGATTTGAGAGCGTCATCGGCTTGGAGAGCTTCTTGATCTTGATCTTGATCTG 400  
 Db 472 TyrLysLysPheGlyLeuIleIleThrSerAlaLeuIleAlaAsnLeuValIleVal 491  
 QY 401 GCAGCTTACAGTATTTG--GATGCGCCACTCACTTTCAGACATCGCTGGATTTCTT 457  
 Db 492 GlyIleMetSerLeuLeuProGlyAlaThrLeuThrMetProGlyIleAlaGlyIleVal 511  
 QY 458 CTTGCTTGGGAGATGCGCGTCATCGGCTTGGAGAGCTTCTTGATCTTGATCTTGATCTG 517  
 Db 512 LeuThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGlnArgIleLysGlnGlu 531  
 QY 518 TTTTATGTCGCAAGCTTAAATAATCTGTAGAAAAAGATATACCAAGCTTTTGA 577  
 Db 532 LeuSerAsnGlyArgThrValGlnGlnAlaIleAspGlnGlyTyrLysGlyAlaPheSer 551  
 QY 578 GCAATTTTGAATCTAATCTGATCTGATGATGATGATGATGATGATGATGATGATGATG 637  
 Db 552 SerIlePheAspAlaAsnValThrThrLeuIleLysValIleIleLeuTyrAlaValGly 571  
 QY 638 ACAGGCGCTTAAAGGCTTGGCTTGAATGATGATGATGATGATGATGATGATGATGATG 697



```

|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
572 ThrGlyAlaIleuYsGlyPheAlaIleThrThrGlyIleGlyAlaThrSerMetPhe 591
Qy 698 ACGGCTCTTTTCATGACTAAA 718
Db 592 ThrAlaIleValGlyThrArg 598
RESULT 45
Q9WZM4 PRELIMINARY; PRT; 465 AA.
AC Q9WZM4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-export membrane protein SEC2, putative.
GN TW0860.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Swinn M.L., Dodson R.J.,
Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Uterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001752; AAD35942.1; -.
DR TIGR; TM0860; -.
DR InterPro; IPR001036; Acrflavin res.
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF; 1.
DR PRINTS; PRO0702; ACRIFLAVINRP.
DR TIGRFAMs; TIGR00916; 2A0604e01; 1.
DR TIGRFAMs; TIGR01129; secD; 1.
KW Complete proteome.
SQ SEQUENCE 465 AA; 50931 MW; 6CE9545B3EFD97F9 CRC64;

```

## Alignment Scores:

```

Pred. No.: 1.23e-21 Length: 465
Score: 355.50 Matches: 78
Percent Similarity: 56.84% Conservative: 55
Best Local Similarity: 33.33% Mismatches: 80
Query Match: 10.73% Indels: 21
DB: 16 Gaps: 5

```

US-09-868-987-1 (1-1864) x Q9WZM4 (1-465)

```

Qy 47 GGACAAATATTCGAAACCGTGGATGCGGTATGCTGTAGTGTGATGACGGTTATATG--- 103
Db 219 GlyThrTyAspProLysLys-----ArgLeuAlaIleValLeuAspValValGln 236
Qy 104 GTCAGCAGCCCTATTTTAAAGCTCCCATGAAATCATGCCAGTGTCTCAGGAAATTT 163
Db 237 PheValGlyGlnValValAlaIleIleThrAspGlyLysAlaGluIleThrGlyAsnPhe 256
Qy 164 ACCCACCCTGAAGTCAGCAACTCGCTCAGATTATAAATCTGGAGCGATG----- 214
Db 257 SerLeuGluGluAlaLysGlnLeuAlaIleLeuArgSerGlyAlaLeuProAlaArg 276
Qy 215 -----TCTTTTGTCCCGAGGTTCTCAGTGAAGACGATCTCTTCT 256
Db 277 LeuValLysThrSerSerGlyTyrValAlaProLeuLeuGlyArgAspValIleAspAla 296
Qy 257 GATCTTGGGAAAAACAATGTACACAGGCATTATCTCAGCATGCTGTGGCTTGGCAATG 316
Db 297 SerLeu-----LysAlaGlyIleIle-----GlyLeuIleLeu 307

```

```

Qy 317 CTATTGCTTTTGATGACGGTATATATATAGATTGGAGCGCTCATCGCTTCGGGAGCTGTT 376
Db 308 ValLeuValTyMetIleIleTyTyArgThrMetGlyIleValAlaAspLeuAlaLeu 327
Qy 377 CTCTGAACTCTTTTGTCTATCTGGCAGCTCTACAGTATTGGATGGCCACTCACCTTG 436
Db 328 IleTyAsnThrValLeuLeuLeuGlyValMetAlaAlaGlyLysPheIleLeuThrLeu 347
Qy 437 TCAGGACTCGCTGGGATTGTCTTCTGATGGGATGGCGCTAGATGCAAAATGTTCTTGT 496
Db 348 ProGlyIleAlaGlyIleIleLeuThrIleGlyThrThrValAspGlyAsnValIleIle 367
Qy 497 TTCGAAAGATCCGAGAGGAATTTTATTCTCTCAAGTCTTAAATAATCTGTAGAAAA 556
Db 368 TyrGluArgIleLysGluGluMetArgLeuGlyLysProValLysThrSerIleAlaAla 387
Qy 557 GGATATACCAAGGCTTTTGGAGCGCATTTTGTATCTAACTTGAATGACTACAGTATTGGCCTCA 616
Db 388 GlyPheAspArgSerLeuSerThrIleLeuAspAlaAsnIleThrThrIleLeuThrGly 407
Qy 617 GCATCTCTTTTCTCTCTAGATACAGGCGCTTATTAAAGGTTTGTCTTGACATTGATTTTA 676
Db 408 LeuIleLeuTyTyPheGlyThrGlyThrIleLysGlyPheAlaIleThrLeuIleIle 427
Qy 677 GGAATTTTCTCTCAATGTTTACGGCTCTTTTCATGACTAAA 718
Db 428 GlyValLeuGlySerIlePheValAsnLeuValPheSerArg 441

```

Search completed: December 17, 2002, 12:38:43  
Job time : 109.5 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

.Run on: December 17, 2002, 09:09:46 ; Search time 64 Seconds  
(without alignments)  
8931.961 Million cell updates/sec

Title: US-09-868-987-1  
Perfect score: 1864  
Sequence: 1 atggaattccgcataattgtc.....atgacaaattcagataatgc 1864

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 1533381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	1.0	2972	2	US-08-720-484A-3
C 2	19	1.0	2972	3	US-08-953-823A-3
C 3	19	1.0	2972	4	US-09-398-239-3
C 4	18	1.0	378	4	US-09-134-001C-182
C 5	18	1.0	1315	2	US-08-343-101A-8
C 6	18	1.0	1315	3	US-09-183-688-8
C 7	18	1.0	2167	4	US-09-325-932A-102
C 8	18	1.0	2259	1	US-08-420-235B-20
C 9	18	1.0	2259	3	US-08-793-624-20
C 10	18	1.0	2259	5	PCT-US95-10194-20
C 11	18	1.0	3715	4	US-09-085-199B-44
C 12	18	1.0	4796	4	US-09-085-199B-3
C 13	18	1.0	20710	1	US-08-420-235B-1
C 14	18	1.0	20710	3	US-08-793-624-1
C 15	18	1.0	20710	5	PCT-US95-10194-1
C 16	18	1.0	28473	4	US-08-961-527-83
C 17	18	1.0	35100	2	US-08-770-379-18
C 18	18	1.0	35100	4	US-08-757-669A-18
C 19	18	1.0	35100	4	US-09-230-371A-18
C 20	18	1.0	36159	4	US-09-749-588-3
C 21	18	1.0	40138	3	US-09-090-753-12
C 22	17	0.9	800	1	US-08-756-771-2
C 23	17	0.9	800	2	US-09-096-571-2
C 24	17	0.9	800	4	US-09-309-320-2
C 25	25	0.9	926	3	US-08-617-860B-21
C 26	17	0.9	1137	4	US-09-045-583-6
C 27	17	0.9	1137	4	US-09-534-185-6
C 28	17	0.9	1664	4	US-09-045-583-4
C 29	17	0.9	1664	4	US-09-534-185-4
C 30	17	0.9	1692	4	US-09-134-001C-79
C 31	17	0.9	1789	4	US-08-936-165A-116
C 32	17	0.9	2114	4	US-09-484-970B-165
C 33	17	0.9	2900	4	US-03-066-047-17
C 34	17	0.9	3854	2	US-08-720-484A-1
C 35	17	0.9	3854	3	US-08-953-823A-1
C 36	17	0.9	3854	4	US-09-398-239-1
C 37	17	0.9	6089	4	US-09-221-017B-8
C 38	17	0.9	7091	4	US-08-975-762-46
C 39	17	0.9	7091	4	US-09-295-028-46
C 40	17	0.9	7091	4	US-03-106-582-46
C 41	17	0.9	7091	4	US-08-961-527-189
C 42	17	0.9	9711	4	US-08-961-527-167
C 43	17	0.9	11785	1	US-08-038-768A-4
C 44	17	0.9	11785	2	US-08-416-603-3
C 45	17	0.9	13158	2	US-08-687-080-105
C 46	17	0.9	22846	2	US-08-469-461-3
C 47	17	0.9	22846	3	US-07-890-609-3
C 48	17	0.9	44453	4	US-09-146-053-5
C 49	17	0.9	99500	4	US-09-798-096-10
C 50	17	0.9	112132	4	US-09-741-150-3
C 51	17	0.9	4411529	4	US-09-103-840A-1
C 52	16	0.9	39	4	US-09-124-541-14
C 53	16	0.9	39	4	US-09-124-541-16
C 54	16	0.9	45	2	US-08-467-603-99
C 55	16	0.9	45	2	US-08-466-793-99
C 56	16	0.9	45	2	US-08-491-861A-99
C 57	16	0.9	71	2	US-08-894-578-191
C 58	16	0.9	71	2	US-08-894-578-215
C 59	16	0.9	366	3	US-08-881-771A-7
C 60	16	0.9	496	2	US-08-967-101-80
C 61	16	0.9	496	2	US-08-592-541-80
C 62	16	0.9	496	3	US-09-124-698-80
C 63	16	0.9	496	4	US-03-127-480-80
C 64	16	0.9	496	4	US-08-436-841C-80
C 65	16	0.9	496	4	US-09-134-523-80
C 66	16	0.9	502	2	US-08-967-101-88
C 67	16	0.9	502	2	US-08-592-541-88
C 68	16	0.9	502	3	US-09-124-698-88
C 69	16	0.9	502	4	US-09-127-480-88
C 70	16	0.9	502	4	US-08-436-841C-88
C 71	16	0.9	502	4	US-09-124-523-88
C 72	16	0.9	585	4	US-09-134-001C-1184
C 73	16	0.9	731	4	US-08-936-165A-170
C 74	16	0.9	763	4	US-09-004-838-67
C 75	16	0.9	774	2	US-08-467-603-71
C 76	16	0.9	774	2	US-08-466-793-71
C 77	16	0.9	774	2	US-08-491-861A-71
C 78	16	0.9	776	1	US-08-358-171-19
C 79	16	0.9	776	3	US-09-090-947-19
C 80	16	0.9	791	2	US-08-467-603-67
C 81	16	0.9	791	2	US-08-466-793-67
C 82	16	0.9	791	2	US-08-491-861A-67
C 83	16	0.9	793	2	US-08-467-603-69
C 84	16	0.9	793	2	US-08-466-793-69
C 85	16	0.9	793	2	US-08-491-861A-69
C 86	16	0.9	800	4	US-08-928-941D-25
C 87	16	0.9	800	4	US-08-928-941D-26
C 88	16	0.9	800	4	US-09-280-590A-25
C 89	16	0.9	800	4	US-09-280-590A-26
C 90	16	0.9	850	4	US-08-928-941D-27
C 91	16	0.9	850	4	US-09-280-590A-27
C 92	16	0.9	809	4	US-08-928-941D-19
C 93	16	0.9	909	4	US-09-280-590A-19
C 94	16	0.9	920	4	US-09-457-046B-1
C 95	16	0.9	982	3	US-08-973-068-28
C 96	16	0.9	998	2	US-08-467-603-73
C 97	16	0.9	998	2	US-08-466-793-73
C 98	16	0.9	998	2	US-08-431-861A-73
C 99	16	0.9	1001	4	US-09-641-638-380
C 100	16	0.9	1001	4	US-09-641-638-381

101	16	0.9	1009	4	US-09-602-877A-107	Sequence 107, App	C 174	16	0.9	3233	3	US-08-755-587-43	Sequence 43, Appl
C 102	16	0.9	1096	3	US-08-881-771A-5	Sequence 5, Appl	C 175	16	0.9	3284	6	5258288-3	Patent No. 5258288
C 103	16	0.9	1101	3	US-08-945-056-5	Sequence 5, Appl	C 176	16	0.9	3290	1	US-07-985-691-1	Sequence 1, Appl
C 104	16	0.9	1103	3	US-08-202-186-16	Sequence 16, Appl	C 177	16	0.9	3290	1	US-08-436-804-1	Sequence 1, Appl
C 105	16	0.9	1104	1	US-08-202-186-17	Sequence 17, Appl	C 178	16	0.9	3290	1	US-08-267-387-1	Sequence 1, Appl
C 106	16	0.9	1109	1	US-08-202-186-13	Sequence 13, Appl	C 179	16	0.9	3680	1	US-08-759-848-1	Sequence 1, Appl
C 107	16	0.9	1110	1	US-08-202-186-11	Sequence 11, Appl	C 180	16	0.9	3680	5	PCT-US95-09383-1	Sequence 1, Appl
C 108	16	0.9	1110	1	US-08-202-186-14	Sequence 14, Appl	C 181	16	0.9	3695	4	US-08-913-159-1	Sequence 1, Appl
C 109	16	0.9	1111	1	US-08-202-186-9	Sequence 9, Appl	C 182	16	0.9	3695	4	US-08-913-159-3	Sequence 3, Appl
C 110	16	0.9	1111	1	US-08-202-186-10	Sequence 10, Appl	C 183	16	0.9	3695	4	US-08-913-159-7	Sequence 7, Appl
C 111	16	0.9	1111	1	US-08-202-186-12	Sequence 12, Appl	C 184	16	0.9	3767	4	US-08-928-941D-28	Sequence 28, Appl
C 112	16	0.9	1111	1	US-08-202-186-24	Sequence 24, Appl	C 185	16	0.9	3767	4	US-08-928-941D-30	Sequence 30, Appl
C 113	16	0.9	1166	1	US-08-121-063-7	Sequence 7, Appl	C 186	16	0.9	3767	4	US-09-280-590A-28	Sequence 28, Appl
C 114	16	0.9	1236	4	US-09-134-001C-284	Sequence 284, App	C 187	16	0.9	3767	4	US-09-280-590A-30	Sequence 30, Appl
C 115	16	0.9	1261	3	US-08-475-742-16	Sequence 16, Appl	C 188	16	0.9	3800	4	US-08-965-762-13	Sequence 13, Appl
C 116	16	0.9	1317	4	US-09-457-046B-57	Sequence 57, Appl	C 189	16	0.9	3800	4	US-09-911-927-13	Sequence 13, Appl
C 117	16	0.9	1347	4	US-09-134-001C-1309	Sequence 1309, Ap	C 190	16	0.9	3800	4	US-09-911-927-15	Sequence 15, Appl
C 118	16	0.9	1462	4	US-08-975-215-3	Sequence 3, Appl	C 191	16	0.9	3800	4	US-09-911-882-13	Sequence 13, Appl
C 119	16	0.9	1462	4	US-09-776-271-3	Sequence 3, Appl	C 192	16	0.9	3800	4	US-09-911-882-15	Sequence 15, Appl
C 120	16	0.9	1510	1	US-08-248-466B-3	Sequence 3, Appl	C 193	16	0.9	3956	2	US-08-966-388-9	Sequence 9, Appl
C 121	16	0.9	1532	1	US-08-248-466B-11	Sequence 11, Appl	C 194	16	0.9	3956	3	US-09-188-403-9	Sequence 9, Appl
C 122	16	0.9	1641	2	US-08-705-771-9	Sequence 9, Appl	C 195	16	0.9	3956	3	US-09-188-404-9	Sequence 9, Appl
C 123	16	0.9	1726	2	US-08-967-101-9	Sequence 9, Appl	C 196	16	0.9	3956	3	US-09-188-404-9	Sequence 9, Appl
C 124	16	0.9	1726	2	US-08-967-101-152	Sequence 152, App	C 197	16	0.9	4328	1	US-08-322-742-12	Sequence 12, Appl
C 125	16	0.9	1726	2	US-08-592-541-9	Sequence 9, Appl	C 198	16	0.9	4328	1	US-08-322-742-15	Sequence 15, Appl
C 126	16	0.9	1726	2	US-08-592-541-152	Sequence 152, App	C 199	16	0.9	4651	2	US-08-449-644-4	Sequence 4, Appl
C 127	16	0.9	1726	3	US-09-124-698-9	Sequence 9, Appl	C 200	16	0.9	4651	2	US-08-087-244A-4	Sequence 4, Appl
C 128	16	0.9	1726	3	US-09-124-698-152	Sequence 152, App	C 201	16	0.9	5111	4	US-09-004-838-118	Sequence 118, App
C 129	16	0.9	1726	4	US-09-127-480-9	Sequence 9, Appl	C 202	16	0.9	5111	4	US-09-178-973B-17	Sequence 17, Appl
C 130	16	0.9	1726	4	US-09-127-480-152	Sequence 152, App	C 203	16	0.9	5935	4	US-09-419-568F-29	Sequence 29, Appl
C 131	16	0.9	1726	4	US-08-496-841C-9	Sequence 9, Appl	C 204	16	0.9	5935	4	US-09-354-243B-29	Sequence 29, Appl
C 132	16	0.9	1726	4	US-08-496-841C-152	Sequence 152, App	C 205	16	0.9	6811	3	US-08-651-472-67	Sequence 67, Appl
C 133	16	0.9	1726	4	US-09-124-523-9	Sequence 9, Appl	C 206	16	0.9	6811	4	US-08-358-928-67	Sequence 67, Appl
C 134	16	0.9	1726	4	US-09-124-523-152	Sequence 152, App	C 207	16	0.9	7176	4	US-09-221-017B-626	Sequence 626, App
C 135	16	0.9	1727	3	US-08-888-077A-7	Sequence 7, Appl	C 208	16	0.9	8787	1	US-08-202-176-3	Sequence 3, Appl
C 136	16	0.9	1737	4	US-09-397-787-59	Sequence 59, Appl	C 209	16	0.9	9707	4	US-08-961-527-164	Sequence 164, App
C 137	16	0.9	1749	4	US-09-134-001C-1893	Sequence 1893, Ap	C 210	16	0.9	10007	4	US-09-410-466-13	Sequence 13, Appl
C 138	16	0.9	1854	4	US-09-134-001C-1175	Sequence 1175, Ap	C 211	16	0.9	11580	4	US-09-334-220-4	Sequence 4, Appl
C 139	16	0.9	1866	1	US-08-403-388-1	Sequence 1, Appl	C 212	16	0.9	12225	2	US-08-822-445-11	Sequence 11, Appl
C 140	16	0.9	1866	3	US-08-558-578-1	Sequence 1, Appl	C 213	16	0.9	12225	2	US-08-822-445-11	Sequence 11, Appl
C 141	16	0.9	1866	4	US-08-446-111D-1	Sequence 1, Appl	C 214	16	0.9	12225	2	US-09-396-540-11	Sequence 11, Appl
C 142	16	0.9	1866	4	US-09-056-105-19	Sequence 19, Appl	C 215	16	0.9	12616	2	US-08-822-445-9	Sequence 9, Appl
C 143	16	0.9	1866	4	US-09-443-077-1	Sequence 7, Appl	C 216	16	0.9	12616	2	US-08-822-445-9	Sequence 9, Appl
C 144	16	0.9	2007	4	US-09-052-089A-7	Sequence 7, Appl	C 217	16	0.9	14602	1	US-08-746-682A-1	Sequence 1, Appl
C 145	16	0.9	2031	3	US-08-262-220-13	Sequence 13, Appl	C 218	16	0.9	14602	1	US-08-597-236-1	Sequence 1, Appl
C 146	16	0.9	2031	3	US-08-471-733-13	Sequence 13, Appl	C 219	16	0.9	20165	4	US-09-609-816-7	Sequence 7, Appl
C 147	16	0.9	2031	3	US-08-468-878-13	Sequence 13, Appl	C 220	16	0.9	40328	3	US-08-742-189-102	Sequence 102, App
C 148	16	0.9	2031	3	US-08-468-878-13	Sequence 13, Appl	C 221	16	0.9	40328	3	US-08-846-111D-15	Sequence 15, Appl
C 149	16	0.9	2031	3	US-08-468-878-13	Sequence 13, Appl	C 222	16	0.9	40352	4	US-09-443-077-15	Sequence 15, Appl
C 150	16	0.9	2031	3	US-08-468-878-13	Sequence 13, Appl	C 223	16	0.9	55827	4	US-09-813-133A-3	Sequence 3, Appl
C 151	16	0.9	2031	3	US-08-468-878-13	Sequence 13, Appl	C 224	16	0.9	98844	4	US-09-791-211-10	Sequence 10, Appl
C 152	16	0.9	2075	3	US-08-750-494-3	Sequence 3, Appl	C 225	16	0.9	176373	3	US-09-128-155-17	Sequence 17, Appl
C 153	16	0.9	2136	4	US-09-134-001C-2623	Sequence 2623, Ap	C 226	16	0.9	29	2	US-08-531-556-99	Sequence 99, Appl
C 154	16	0.9	2264	3	US-08-262-220-5	Sequence 5, Appl	C 227	16	0.9	29	2	US-08-690-734A-30	Sequence 30, Appl
C 155	16	0.9	2264	3	US-08-471-733-5	Sequence 5, Appl	C 228	16	0.9	29	2	US-08-742-185-30	Sequence 30, Appl
C 156	16	0.9	2264	3	US-08-468-878-5	Sequence 5, Appl	C 229	16	0.9	39	3	US-08-766-433-96	Sequence 96, Appl
C 157	16	0.9	2264	3	US-08-468-878-5	Sequence 5, Appl	C 230	16	0.9	37	4	US-09-082-649B-8	Sequence 8, Appl
C 158	16	0.9	2325	2	US-08-966-388-5	Sequence 5, Appl	C 231	16	0.9	251	4	US-08-659-473-1	Sequence 1, Appl
C 159	16	0.9	2325	2	US-09-188-403-5	Sequence 5, Appl	C 232	16	0.9	251	4	US-09-385-989-103	Sequence 103, App
C 160	16	0.9	2325	3	US-09-188-404-5	Sequence 5, Appl	C 233	16	0.9	251	4	US-09-222-575-101	Sequence 101, App
C 161	16	0.9	2325	3	US-09-281-359-5	Sequence 5, Appl	C 234	16	0.9	357	1	US-08-324-799-4	Sequence 4, Appl
C 162	16	0.9	2397	6	5258288-2	Patent No. 5258288	C 235	16	0.9	357	1	US-08-192-102-4	Sequence 4, Appl
C 163	16	0.9	2547	3	US-08-262-220-7	Sequence 7, Appl	C 236	16	0.9	357	2	US-08-192-861A-4	Sequence 4, Appl
C 164	16	0.9	2547	3	US-08-471-733-7	Sequence 7, Appl	C 237	16	0.9	357	4	US-09-133-119-4	Sequence 4, Appl
C 165	16	0.9	2547	3	US-08-468-878-7	Sequence 7, Appl	C 238	16	0.9	357	4	US-08-192-093A-4	Sequence 4, Appl
C 166	16	0.9	2547	3	US-08-750-494-7	Sequence 7, Appl	C 239	16	0.9	361	1	US-08-471-780C-106	Sequence 106, App
C 167	16	0.9	2568	6	5196526-3	Patent No. 5196526	C 240	16	0.9	361	1	US-08-467-282B-106	Sequence 106, App
C 168	16	0.9	2903	4	US-08-928-941D-2	Sequence 2, Appl	C 241	16	0.9	361	2	US-08-471-710C-106	Sequence 106, App
C 169	16	0.9	2903	4	US-08-928-941D-3	Sequence 3, Appl	C 242	16	0.9	361	2	US-08-466-710C-106	Sequence 106, App
C 170	16	0.9	2903	4	US-09-280-590A-2	Sequence 2, Appl	C 243	16	0.9	361	3	US-08-471-284A-106	Sequence 106, App
C 171	16	0.9	2903	4	US-09-280-590A-3	Sequence 3, Appl	C 244	16	0.9	384	4	US-08-468-739E-106	Sequence 106, App
C 172	16	0.9	2945	6	5196526-2	Patent No. 5196526	C 245	16	0.9	384	4	US-09-134-001A-99	Sequence 99, Appl
C 173	16	0.9	3030	2	US-08-660-326-24	Sequence 24, Appl	C 246	16	0.9	389	4	US-09-359-301A-26	Sequence 26, Appl

C 247	15	0.8	400	4	US-09-359-301A-22	Sequence 22, Appl	320	15	0.8	1431	4	US-09-316-083-2	Sequence 2, Appl
C 248	15	0.8	400	4	US-09-370-838-143	Sequence 143, App	C 321	15	0.8	1432	5	FCT-US94-12912-4	Sequence 4, Appl
C 249	15	0.8	400	4	US-09-370-838-147	Sequence 147, App	C 322	15	0.8	1455	4	US-09-149-476-272	Sequence 272, App
C 250	15	0.8	404	4	US-09-359-301A-21	Sequence 21, Appl	323	15	0.8	1458	4	US-09-334-938-9	Sequence 9, Appl
C 251	15	0.8	423	1	US-08-470-179-146	Sequence 146, App	324	15	0.8	1539	2	US-08-899-811-22	Sequence 22, Appl
C 252	15	0.8	437	4	US-08-943-731-147	Sequence 147, App	325	15	0.8	1542	2	US-09-122-230-8	Sequence 8, Appl
C 253	15	0.8	441	4	US-08-397-787-275	Sequence 275, App	C 326	15	0.8	1617	4	US-09-265-013-2	Sequence 2, Appl
C 254	15	0.8	442	4	US-08-227-357-132	Sequence 132, App	C 327	15	0.8	1638	4	US-09-149-476-271	Sequence 271, App
C 255	15	0.8	499	4	US-08-642-274D-34	Sequence 34, Appl	C 328	15	0.8	1644	4	US-09-149-476-116	Sequence 116, App
C 256	15	0.8	499	4	US-08-952-014C-34	Sequence 34, Appl	C 329	15	0.8	1644	4	US-09-398-395A-3	Sequence 3, Appl
C 257	15	0.8	528	4	US-09-452-239-33	Sequence 33, Appl	C 330	15	0.8	1644	4	US-09-398-395A-7	Sequence 7, Appl
C 258	15	0.8	550	4	US-09-359-301A-23	Sequence 23, Appl	C 331	15	0.8	1644	4	US-09-398-395A-5	Sequence 5, Appl
C 259	15	0.8	563	4	US-09-221-017B-645	Sequence 645, App	C 332	15	0.8	1644	4	US-09-398-395A-9	Sequence 9, Appl
C 260	15	0.8	566	4	US-09-328-111-455	Sequence 455, App	C 333	15	0.8	1644	4	US-09-398-395A-11	Sequence 11, Appl
C 261	15	0.8	609	4	US-09-347-833-9	Sequence 9, Appl	C 334	15	0.8	1650	1	US-08-459-100A-1	Sequence 1, Appl
C 262	15	0.8	614	4	US-09-325-932A-31	Sequence 31, Appl	C 335	15	0.8	1650	5	FCT-US94-09589-1	Sequence 1, Appl
C 263	15	0.8	641	4	US-09-221-017B-177	Sequence 177, App	336	15	0.8	1666	1	US-08-848-932-1	Sequence 1, Appl
C 264	15	0.8	644	4	US-09-328-111-117	Sequence 117, App	337	15	0.8	1666	2	US-09-008-180-1	Sequence 1, Appl
C 265	15	0.8	655	3	US-08-714-918-100	Sequence 100, App	C 338	15	0.8	1716	3	US-08-656-034-9	Sequence 9, Appl
C 266	15	0.8	655	4	US-09-265-315-100	Sequence 100, App	339	15	0.8	1762	4	US-09-443-184-35	Sequence 35, Appl
C 267	15	0.8	655	4	US-09-265-315-100	Sequence 100, App	C 340	15	0.8	1778	1	US-08-416-870C-3	Sequence 3, Appl
C 268	15	0.8	655	4	US-09-265-417-100	Sequence 100, App	341	15	0.8	1796	4	US-09-255-392-1	Sequence 1, Appl
C 269	15	0.8	675	4	US-09-221-017B-612	Sequence 612, App	C 342	15	0.8	1803	4	US-09-134-001C-2456	Sequence 2456, App
C 270	15	0.8	688	4	US-08-998-416-969	Sequence 969, App	C 343	15	0.8	1815	4	US-09-586-935-2	Sequence 2, Appl
C 271	15	0.8	705	4	US-09-219-337-1	Sequence 1, Appl	C 344	15	0.8	1817	2	US-08-870-518-8	Sequence 8, Appl
C 272	15	0.8	749	4	US-09-166-350-8	Sequence 8, Appl	345	15	0.8	1864	2	US-08-759-581B-19	Sequence 19, Appl
C 273	15	0.8	751	4	US-09-071-035-59	Sequence 59, Appl	346	15	0.8	1864	4	US-09-304-711-19	Sequence 19, Appl
C 274	15	0.8	773	4	US-09-359-301A-20	Sequence 20, Appl	347	15	0.8	1864	4	US-09-173-281-19	Sequence 19, Appl
C 275	15	0.8	789	1	US-08-181-335B-3	Sequence 3, Appl	348	15	0.8	1896	4	US-09-343-011B-4	Sequence 4, Appl
C 276	15	0.8	789	1	US-08-181-335B-5	Sequence 5, Appl	C 349	15	0.8	1909	2	US-09-122-230-6	Sequence 6, Appl
C 277	15	0.8	789	1	US-08-181-335B-6	Sequence 6, Appl	C 350	15	0.8	1914	1	US-08-350-325A-9	Sequence 9, Appl
C 278	15	0.8	789	5	PCT-US95-00129-3	Sequence 3, Appl	C 351	15	0.8	1914	5	FCT-US94-03856-9	Sequence 9, Appl
C 279	15	0.8	789	5	PCT-US95-00129-5	Sequence 5, Appl	C 352	15	0.8	1919	1	US-07-991-587A-1	Sequence 1, Appl
C 280	15	0.8	789	5	PCT-US95-00129-6	Sequence 6, Appl	C 353	15	0.8	1919	1	US-08-309-985-1	Sequence 1, Appl
C 281	15	0.8	792	4	US-09-134-001C-2394	Sequence 2394, App	354	15	0.8	1977	4	US-09-626-589-5	Sequence 5, Appl
C 282	15	0.8	808	3	US-08-765-332-131	Sequence 131, App	355	15	0.8	1996	3	US-08-714-918-44	Sequence 44, Appl
C 283	15	0.8	808	3	US-08-765-332-132	Sequence 132, App	356	15	0.8	1996	4	US-09-285-315-44	Sequence 44, Appl
C 284	15	0.8	808	4	US-09-448-894-131	Sequence 131, App	357	15	0.8	1996	4	US-09-265-315-44	Sequence 44, Appl
C 285	15	0.8	808	4	US-09-448-894-132	Sequence 132, App	358	15	0.8	1996	4	US-09-266-415-44	Sequence 44, Appl
C 286	15	0.8	809	3	US-08-765-332-154	Sequence 154, App	359	15	0.8	2005	2	US-08-484-200-1	Sequence 1, Appl
C 287	15	0.8	809	4	US-09-448-894-154	Sequence 154, App	360	15	0.8	2005	2	US-08-477-493-1	Sequence 1, Appl
C 288	15	0.8	867	4	US-09-071-035-57	Sequence 57, Appl	361	15	0.8	2005	3	US-08-465-375-1	Sequence 1, Appl
C 289	15	0.8	912	4	US-09-221-017B-763	Sequence 763, App	362	15	0.8	2005	4	US-09-788-070-1	Sequence 1, Appl
C 290	15	0.8	969	4	US-09-080-205-7	Sequence 7, Appl	C 363	15	0.8	2016	4	US-09-132-118-1	Sequence 1, Appl
C 291	15	0.8	974	1	US-08-848-932-3	Sequence 3, Appl	C 364	15	0.8	2060	2	US-08-417-210A-142	Sequence 142, App
C 292	15	0.8	974	2	US-09-008-180-3	Sequence 3, Appl	C 365	15	0.8	2065	1	US-08-968-751-1	Sequence 1, Appl
C 293	15	0.8	998	2	US-07-885-089B-5	Sequence 5, Appl	366	15	0.8	2106	3	US-08-613-009A-4	Sequence 4, Appl
C 294	15	0.8	1001	4	US-09-641-638-376	Sequence 376, App	367	15	0.8	2106	4	US-08-778-570B-4	Sequence 4, Appl
C 295	15	0.8	1001	4	US-09-641-638-377	Sequence 377, App	368	15	0.8	2106	4	US-09-059-584-4	Sequence 4, Appl
C 296	15	0.8	1035	4	US-09-073-898-139	Sequence 139, App	C 369	15	0.8	2137	1	US-08-444-005-16	Sequence 16, Appl
C 297	15	0.8	1101	3	US-08-945-056-7	Sequence 7, Appl	370	15	0.8	2151	4	US-09-626-589-4	Sequence 4, Appl
C 298	15	0.8	1101	4	US-09-638-715-1	Sequence 1, Appl	C 371	15	0.8	2160	3	US-08-656-034-1	Sequence 1, Appl
C 299	15	0.8	1101	4	US-09-638-715-3	Sequence 3, Appl	C 372	15	0.8	2162	2	US-08-870-518-5	Sequence 5, Appl
C 300	15	0.8	1147	1	US-08-665-716-1	Sequence 1, Appl	373	15	0.8	2175	2	US-08-389-564B-4	Sequence 4, Appl
C 301	15	0.8	1200	3	US-08-867-381A-1	Sequence 1, Appl	374	15	0.8	2175	3	US-08-466-047B-4	Sequence 4, Appl
C 302	15	0.8	1200	4	US-09-521-144-1	Sequence 1, Appl	375	15	0.8	2247	3	US-08-613-009A-3	Sequence 3, Appl
C 303	15	0.8	1202	3	US-09-058-489-29	Sequence 29, Appl	376	15	0.8	2247	4	US-08-778-570B-3	Sequence 3, Appl
C 304	15	0.8	1207	3	US-09-166-412-1	Sequence 1, Appl	377	15	0.8	2247	4	US-09-059-584-3	Sequence 3, Appl
C 305	15	0.8	1207	4	US-08-731-320B-1	Sequence 1, Appl	C 378	15	0.8	2259	1	US-08-119-773-35	Sequence 35, Appl
C 306	15	0.8	1212	1	US-08-282-581-1	Sequence 1, Appl	379	15	0.8	2271	4	US-09-134-001C-716	Sequence 716, App
C 307	15	0.8	1212	1	US-08-550-544-1	Sequence 1, Appl	380	15	0.8	2271	2	US-08-788-539A-1	Sequence 1, Appl
C 308	15	0.8	1253	2	US-08-786-606-6	Sequence 6, Appl	381	15	0.8	2362	1	US-08-920-812-5	Sequence 5, Appl
C 309	15	0.8	1260	4	US-09-166-350-22	Sequence 22, Appl	382	15	0.8	2362	1	US-08-920-827-5	Sequence 5, Appl
C 310	15	0.8	1263	4	US-09-222-938A-53	Sequence 53, Appl	383	15	0.8	2362	1	US-08-921-177-5	Sequence 5, Appl
C 311	15	0.8	1304	2	US-08-766-439-37	Sequence 37, Appl	384	15	0.8	2362	2	US-08-362-577C-5	Sequence 5, Appl
C 312	15	0.8	1304	2	US-08-766-439-38	Sequence 38, Appl	385	15	0.8	2362	2	US-08-920-828-5	Sequence 5, Appl
C 313	15	0.8	1326	4	US-09-138-448-28	Sequence 28, Appl	386	15	0.8	2371	2	US-09-192-466-1	Sequence 1, Appl
C 314	15	0.8	1346	2	US-09-123-851-2	Sequence 2, Appl	387	15	0.8	2371	3	US-09-132-659-1	Sequence 1, Appl
C 315	15	0.8	1346	2	US-08-728-520-2	Sequence 2, Appl	388	15	0.8	2381	2	US-08-736-770-4	Sequence 4, Appl
C 316	15	0.8	1351	4	US-09-221-017B-236	Sequence 236, App	389	15	0.8	2399	4	US-09-385-801-1	Sequence 1, Appl
C 317	15	0.8	1391	2	US-08-950-168-2	Sequence 2, Appl	C 390	15	0.8	2419	1	US-08-428-415-1	Sequence 1, Appl
C 318	15	0.8	1391	4	US-09-365-705-2	Sequence 2, Appl	C 391	15	0.8	2419	1	US-08-379-685-1	Sequence 1, Appl
C 319	15	0.8	1404	4	US-09-257-179-34	Sequence 34, Appl	C 392	15	0.8	2419	2	US-08-854-029-1	Sequence 1, Appl



C 539	15	0.8	18318	2	US-08-926-922-6	Sequence 6, Appli	612	14	0.8	174	6	5190931-2	Patent No. 5190931
C 540	15	0.8	18318	3	US-09-233-662-6	Sequence 6, Appli	613	14	0.8	180	1	US-08-222-177A-6	Sequence 6, Appli
C 541	15	0.8	18318	4	US-09-527-657-6	Sequence 6, Appli	614	14	0.8	182	1	US-08-266-311-13	Sequence 13, Appli
542	15	0.8	18627	4	US-08-961-527-113	Sequence 113, App	615	14	0.8	182	1	US-08-467-527A-13	Sequence 13, Appli
543	15	0.8	19250	4	US-08-961-527-35	Sequence 35, Appli	616	14	0.8	182	1	US-08-467-527A-13	Sequence 13, Appli
C 544	15	0.8	19446	4	US-08-961-527-51	Sequence 51, Appli	617	14	0.8	184	1	US-08-594-031-38	Sequence 38, Appli
C 545	15	0.8	20986	4	US-08-961-527-54	Sequence 54, Appli	618	14	0.8	193	1	US-08-594-031-38	Sequence 38, Appli
C 546	15	0.8	22306	4	US-09-433-702B-251	Sequence 251, App	619	14	0.8	203	1	US-09-134-001C-1829	Sequence 1829, App
547	15	0.8	24183	4	US-08-943-731-3	Sequence 3, Appli	620	14	0.8	213	4	US-09-134-001C-1829	Sequence 1829, App
548	15	0.8	24417	2	US-08-846-762-1	Sequence 1, Appli	621	14	0.8	223	4	US-09-134-001C-663	Sequence 663, App
C 550	15	0.8	38584	4	US-09-453-702B-50	Sequence 50, Appli	622	14	0.8	234	4	US-08-248-474-70	Sequence 70, Appli
C 551	15	0.8	38844	4	US-09-734-675-3	Sequence 3, Appli	623	14	0.8	238	3	US-08-756-849-70	Sequence 70, Appli
C 552	15	0.8	41708	4	US-09-470-512A-3	Sequence 3, Appli	624	14	0.8	254	4	US-09-468-738A-12	Sequence 12, Appli
C 553	15	0.8	48974	4	US-08-920-432-17	Sequence 17, Appli	625	14	0.8	254	4	US-09-940-019-12	Sequence 12, Appli
553	15	0.8	51259	3	US-08-781-891-209	Sequence 209, App	626	14	0.8	262	4	US-08-944-483-5	Sequence 5, Appli
554	15	0.8	51952	3	US-08-947-823-1	Sequence 1, Appli	627	14	0.8	262	5	PCT-US95-0933A-3	Sequence 3, Appli
555	15	0.8	56516	2	US-08-996-306-1	Sequence 1, Appli	628	14	0.8	267	2	US-08-928-613-3	Sequence 3, Appli
556	15	0.8	56516	4	US-09-338-907-1	Sequence 1, Appli	629	14	0.8	267	3	US-09-193-524-3	Sequence 3, Appli
557	15	0.8	56516	4	US-09-218-207-1	Sequence 1, Appli	630	14	0.8	267	3	US-09-134-001C-2321	Sequence 2321, App
558	15	0.8	56520	4	US-09-338-907-179	Sequence 179, App	631	14	0.8	270	2	US-08-520-678A-30	Sequence 30, Appli
559	15	0.8	56520	4	US-09-218-207-179	Sequence 179, App	632	14	0.8	270	4	US-08-897-126-30	Sequence 30, Appli
560	15	0.8	65042	4	US-09-784-316-3	Sequence 3, Appli	633	14	0.8	286	3	US-08-688-988-12	Sequence 12, Appli
C 561	15	0.8	72928	3	US-09-009-913-1	Sequence 1, Appli	634	14	0.8	301	4	US-09-605-785-291	Sequence 291, App
C 562	15	0.8	81001	4	US-09-750-580-1	Sequence 1, Appli	635	14	0.8	301	4	US-09-439-313-291	Sequence 291, App
C 563	15	0.8	84495	4	US-09-797-906-3	Sequence 3, Appli	636	14	0.8	301	4	US-09-352-616A-291	Sequence 291, App
C 564	15	0.8	87563	4	US-09-453-702B-57	Sequence 57, Appli	637	14	0.8	301	4	US-09-232-149A-291	Sequence 291, App
C 565	15	0.8	112132	4	US-09-741-150-3	Sequence 3, Appli	638	14	0.8	307	4	US-09-230-670C-3	Sequence 3, Appli
C 566	15	0.8	168575	4	US-09-436-290-1	Sequence 1, Appli	639	14	0.8	309	3	US-09-023-221A-5	Sequence 5, Appli
C 567	15	0.8	441529	4	US-09-103-840A-1	Sequence 1, Appli	640	14	0.8	309	3	US-09-023-221A-22	Sequence 22, Appli
568	14	0.8	20	1	US-07-922-723A-52	Sequence 52, Appli	641	14	0.8	309	4	US-09-282-352A-5	Sequence 5, Appli
569	14	0.8	20	1	US-07-799-828C-52	Sequence 52, Appli	642	14	0.8	309	4	US-09-282-352A-22	Sequence 22, Appli
C 570	14	0.8	20	2	US-08-468-352-49	Sequence 49, Appli	643	14	0.8	309	4	US-09-134-001C-2158	Sequence 2158, App
571	14	0.8	20	2	US-07-952-277A-52	Sequence 52, Appli	644	14	0.8	319	4	US-09-071-035-171	Sequence 171, App
C 572	14	0.8	21	3	US-08-957-302A-7	Sequence 7, Appli	645	14	0.8	327	4	US-09-134-001C-2632	Sequence 2632, App
573	14	0.8	21	4	US-09-217-490-21	Sequence 21, Appli	646	14	0.8	332	4	US-09-144-085-6	Sequence 6, Appli
C 574	14	0.8	21	4	US-09-542-403-7	Sequence 7, Appli	647	14	0.8	332	4	US-09-280-116-261	Sequence 261, App
575	14	0.8	25	4	US-09-360-186-8	Sequence 8, Appli	648	14	0.8	340	4	US-09-422-575-155	Sequence 155, App
576	14	0.8	30	1	US-08-484-557C-55	Sequence 55, Appli	649	14	0.8	345	4	US-09-127-946-5	Sequence 5, Appli
577	14	0.8	30	1	US-08-487-426B-55	Sequence 55, Appli	650	14	0.8	346	1	US-08-263-413-13	Sequence 13, Appli
578	14	0.8	30	2	US-08-487-720A-55	Sequence 55, Appli	651	14	0.8	350	1	US-08-263-413-12	Sequence 12, Appli
579	14	0.8	30	4	US-09-258-797-74	Sequence 74, Appli	652	14	0.8	351	1	US-08-190-687B-15	Sequence 15, Appli
580	14	0.8	34	1	US-08-741-881-44	Sequence 44, Appli	653	14	0.8	360	6	5273901-1	Patent No. 5273901
581	14	0.8	34	1	US-08-739-158-44	Sequence 44, Appli	654	14	0.8	378	3	US-09-023-221A-21	Sequence 21, Appli
582	14	0.8	34	2	US-08-739-167-44	Sequence 44, Appli	655	14	0.8	378	4	US-09-282-352A-21	Sequence 21, Appli
583	14	0.8	34	3	US-08-404-796-44	Sequence 44, Appli	656	14	0.8	378	4	US-09-071-035-169	Sequence 169, App
584	14	0.8	34	3	US-08-931-869-44	Sequence 44, Appli	657	14	0.8	384	4	US-09-199-637A-254	Sequence 254, App
585	14	0.8	34	4	US-09-350-399-44	Sequence 44, Appli	658	14	0.8	384	4	US-09-325-932A-26	Sequence 26, Appli
586	14	0.8	34	4	US-09-236-140A-44	Sequence 44, Appli	659	14	0.8	391	4	US-09-359-301A-27	Sequence 27, Appli
587	14	0.8	34	4	US-09-415-784-41	Sequence 41, Appli	660	14	0.8	402	4	US-08-991-789A-143	Sequence 143, App
588	14	0.8	34	4	US-08-944-465A-41	Sequence 41, Appli	661	14	0.8	402	4	US-09-062-451-143	Sequence 143, App
589	14	0.8	34	4	US-09-415-868-41	Sequence 41, Appli	662	14	0.8	402	4	US-09-598-326-143	Sequence 143, App
590	14	0.8	34	4	US-09-415-900-41	Sequence 41, Appli	663	14	0.8	419	4	US-09-247-155-137	Sequence 137, App
591	14	0.8	34	4	US-09-415-900-41	Sequence 41, Appli	664	14	0.8	421	4	US-09-008-697A-9	Sequence 9, Appli
592	14	0.8	35	2	US-08-801-998A-28	Sequence 28, Appli	665	14	0.8	421	4	US-09-404-879A-157	Sequence 157, App
593	14	0.8	38	4	US-08-466-356-2	Sequence 2, Appli	666	14	0.8	429	4	US-09-247-155-73	Sequence 73, Appli
C 594	14	0.8	41	4	US-08-906-930-3	Sequence 3, Appli	667	14	0.8	435	4	US-09-144-367-7	Sequence 7, Appli
595	14	0.8	42	3	US-08-997-758-2	Sequence 2, Appli	668	14	0.8	431	4	US-09-328-111-122	Sequence 122, App
596	14	0.8	42	4	US-09-468-872-8	Sequence 8, Appli	669	14	0.8	438	4	US-09-134-001C-2406	Sequence 2406, App
C 597	14	0.8	44	1	US-08-471-791-37	Sequence 37, Appli	670	14	0.8	438	4	US-08-998-416-195	Sequence 195, App
C 598	14	0.8	44	5	PCT-US91-01746-37	Sequence 37, Appli	671	14	0.8	443	4	US-09-397-787-130	Sequence 130, App
C 599	14	0.8	50	1	US-08-207-901-88	Sequence 88, Appli	672	14	0.8	443	4	US-08-552-369-9	Sequence 9, Appli
C 600	14	0.8	72	1	US-08-232-463-39	Sequence 39, Appli	673	14	0.8	453	4	US-09-134-001C-186	Sequence 186, App
601	14	0.8	78	3	US-08-945-734-55	Sequence 55, Appli	674	14	0.8	460	3	US-08-961-083-223	Sequence 223, App
602	14	0.8	78	4	US-09-258-797-50	Sequence 50, Appli	675	14	0.8	460	4	US-09-280-116-216	Sequence 216, App
603	14	0.8	78	4	US-08-899-241-226	Sequence 226, App	676	14	0.8	468	1	US-08-307-499-47	Sequence 47, Appli
604	14	0.8	78	5	PCT-US96-09451-50	Sequence 50, Appli	677	14	0.8	468	4	US-09-299-268-47	Sequence 47, Appli
C 605	14	0.8	80	4	US-08-899-241-227	Sequence 27, App	678	14	0.8	474	1	US-08-173-209A-1	Sequence 1, Appli
C 606	14	0.8	89	1	US-08-232-463-46	Sequence 46, Appli	679	14	0.8	479	4	US-09-221-017B-652	Sequence 652, App
607	14	0.8	143	2	US-08-835-231-14	Sequence 14, Appli	680	14	0.8	483	4	US-09-392-184-21	Sequence 21, Appli
608	14	0.8	143	4	US-09-108-661-14	Sequence 14, Appli	681	14	0.8	486	2	US-08-506-864A-5	Sequence 5, Appli
C 609	14	0.8	150	6	5182210-5	Patent No. 5182210	682	14	0.8	486	2	US-08-851-968-5	Sequence 5, Appli
610	14	0.8	159	2	US-08-835-231-15	Sequence 15, Appli	683	14	0.8	491	1	US-07-945-288-7	Sequence 7, Appli
611	14	0.8	159	4	US-09-108-661-15	Sequence 15, Appli	684	14	0.8	491	1	US-08-462-831-7	Sequence 7, Appli

C 685	14	0.8	491 1	US-08-461-809-7	Sequence 7, Appl.	758	14	0.8	665 4	US-09-352-616A-196	Sequence 196, App
C 686	14	0.8	491 1	US-08-461-441-7	Sequence 7, Appl.	759	14	0.8	665 4	US-09-232-149A-196	Sequence 196, App
C 687	14	0.8	491 1	US-08-482-142-7	Sequence 7, Appl.	760	14	0.8	669 4	US-08-943-731-90	Sequence 90, Appl
C 688	14	0.8	491 2	US-08-478-572-7	Sequence 7, Appl.	761	14	0.8	673 4	US-09-280-116-93	Sequence 93, Appl
C 689	14	0.8	491 4	US-08-484-296-7	Sequence 7, Appl.	762	14	0.8	675 4	US-08-998-416-1079	Sequence 1079, Ap
C 690	14	0.8	491 5	PCT-US93-08518-7	Sequence 7, Appl.	763	14	0.8	676 4	US-09-221-017B-1080	Sequence 1080, Ap
C 691	14	0.8	495 4	US-09-280-116-153	Sequence 153, Appl	764	14	0.8	679 4	US-09-346-510B-31	Sequence 31, Appl
C 692	14	0.8	498 1	US-08-413-803-28	Sequence 28, Appl	765	14	0.8	685 4	US-09-221-017B-861	Sequence 861, App
C 693	14	0.8	498 1	US-09-134-001C-2494	Sequence 2494, Ap	766	14	0.8	688 4	US-08-998-416-983	Sequence 983, App
C 694	14	0.8	500 6	5190931-1	Patent No. 5190931	767	14	0.8	690 4	US-09-404-879A-321	Sequence 321, App
C 695	14	0.8	503 4	US-09-484-970B-91	Sequence 91, Appl	768	14	0.8	693 4	US-09-187-789-8	Sequence 8, Appl
C 696	14	0.8	504 6	US-09-221-017B-166	Sequence 166, App	769	14	0.8	697 3	US-08-871-722A-16	Sequence 16, Appl
C 697	14	0.8	509 6	5182210-15	Patent No. 5182210	770	14	0.8	697 4	US-09-346-510B-16	Sequence 16, Appl
C 698	14	0.8	517 2	US-08-327-451E-22	Sequence 22, Appl	771	14	0.8	699 4	US-09-091-838-65	Sequence 65, Appl
C 699	14	0.8	517 2	US-08-458-109-22	Sequence 22, Appl	772	14	0.8	706 4	US-08-998-416-1116	Sequence 1116, Ap
C 700	14	0.8	521 4	US-08-858-207A-215	Sequence 215, App	773	14	0.8	706 4	US-09-134-001C-2308	Sequence 2308, Ap
C 701	14	0.8	537 4	US-08-642-274D-50	Sequence 50, Appl	774	14	0.8	710 4	US-08-998-416-629	Sequence 629, App
C 702	14	0.8	537 4	US-08-952-014C-50	Sequence 50, Appl	775	14	0.8	711 4	US-08-998-416-630	Sequence 630, App
C 703	14	0.8	537 4	US-09-134-001C-817	Sequence 817, App	776	14	0.8	722 1	US-07-971-160-3	Sequence 3, Appl
C 704	14	0.8	539 4	US-09-385-982-198	Sequence 198, App	777	14	0.8	722 1	US-08-465-273-3	Sequence 3, Appl
C 705	14	0.8	539 6	5182210-13	Patent No. 5182210	778	14	0.8	722 2	US-09-119-024-3	Sequence 3, Appl
C 706	14	0.8	541 4	US-08-714-918-73	Sequence 73, Appl	779	14	0.8	722 2	US-08-417-226-3	Sequence 3, Appl
C 707	14	0.8	541 4	US-09-265-315-73	Sequence 73, Appl	780	14	0.8	722 2	US-09-180-077-5	Sequence 5, Appl
C 708	14	0.8	541 4	US-09-265-315-73	Sequence 73, Appl	781	14	0.8	722 4	US-09-196-131-3	Sequence 3, Appl
C 709	14	0.8	541 4	US-09-265-315-73	Sequence 73, Appl	782	14	0.8	722 4	US-08-643-732-3	Sequence 3, Appl
C 710	14	0.8	544 5	PCT-US93-08744-1	Sequence 1, Appl	783	14	0.8	729 4	US-09-276-531-43	Sequence 43, Appl
C 711	14	0.8	545 4	US-08-446-935-5	Sequence 5, Appl	784	14	0.8	729 4	US-09-134-001C-1817	Sequence 1817, Ap
C 712	14	0.8	549 1	US-07-991-867B-28	Sequence 28, Appl	785	14	0.8	732 4	US-09-328-111-50	Sequence 50, Appl
C 713	14	0.8	549 1	US-08-107-755A-28	Sequence 28, Appl	786	14	0.8	733 4	US-09-071-035-181	Sequence 191, App
C 714	14	0.8	549 2	US-08-544-332-28	Sequence 28, Appl	787	14	0.8	736 4	US-08-188-582-23	Sequence 23, Appl
C 715	14	0.8	549 4	US-09-091-590A-13	Sequence 13, Appl	788	14	0.8	738 1	US-08-646-715-23	Sequence 23, Appl
C 716	14	0.8	549 4	US-09-370-861A-28	Sequence 28, Appl	789	14	0.8	738 1	US-08-646-715-23	Sequence 23, Appl
C 717	14	0.8	550 4	US-08-632-511A-7	Sequence 7, Appl	790	14	0.8	742 2	US-09-662-250A-14	Sequence 14, Appl
C 718	14	0.8	550 4	US-09-488-200-7	Sequence 7, Appl	791	14	0.8	742 2	US-08-799-464A-15	Sequence 15, Appl
C 719	14	0.8	553 4	US-09-392-184-3	Sequence 3, Appl	792	14	0.8	747 5	PCT-US95-09977-15	Sequence 75, Appl
C 720	14	0.8	553 2	US-08-835-231-3	Sequence 3, Appl	793	14	0.8	750 5	PCT-US95-10964-72	Sequence 72, Appl
C 721	14	0.8	561 2	US-08-835-231-3	Sequence 3, Appl	794	14	0.8	760 3	US-09-023-221A-1	Sequence 1, Appl
C 722	14	0.8	561 4	US-09-108-661-3	Sequence 4, Appl	795	14	0.8	762 4	US-09-282-352A-1	Sequence 1, Appl
C 723	14	0.8	561 4	US-09-108-661-4	Sequence 4, Appl	796	14	0.8	762 4	US-09-134-001C-2598	Sequence 2598, Ap
C 724	14	0.8	563 1	US-08-463-115-73	Sequence 73, Appl	797	14	0.8	765 4	US-08-998-416-215	Sequence 215, App
C 725	14	0.8	563 1	US-08-465-388-73	Sequence 73, Appl	798	14	0.8	770 4	US-09-180-077-10	Sequence 10, Appl
C 726	14	0.8	565 4	US-09-328-111-51	Sequence 51, Appl	799	14	0.8	774 4	US-08-998-416-655	Sequence 655, App
C 727	14	0.8	578 1	US-07-764-731B-1	Sequence 1, Appl	800	14	0.8	777 4	US-09-187-789-4	Sequence 4, Appl
C 728	14	0.8	578 6	5187076-1	Patent No. 5187076	801	14	0.8	777 4	US-09-149-476-52	Sequence 52, Appl
C 729	14	0.8	579 4	US-08-487-429A-9	Sequence 9, Appl	802	14	0.8	800 2	US-08-929-302-3	Sequence 3, Appl
C 730	14	0.8	597 4	US-09-221-017B-705	Sequence 705, App	803	14	0.8	800 2	US-09-038-014-3	Sequence 3, Appl
C 731	14	0.8	602 4	US-09-385-982-269	Sequence 269, App	804	14	0.8	800 3	US-08-357-497-1	Sequence 7, Appl
C 732	14	0.8	607 1	US-08-240-748A-1	Sequence 1, Appl	805	14	0.8	807 4	US-08-861-774B-67	Sequence 67, Appl
C 733	14	0.8	616 1	US-08-240-748A-4	Sequence 4, Appl	806	14	0.8	809 4	US-09-134-001C-2360	Sequence 2360, Ap
C 734	14	0.8	616 1	US-08-240-748A-6	Sequence 6, Appl	807	14	0.8	810 4	US-09-221-017B-900	Sequence 900, App
C 735	14	0.8	618 1	US-08-361-467B-12	Sequence 12, Appl	808	14	0.8	818 4	US-09-071-035-189	Sequence 189, App
C 736	14	0.8	618 1	US-08-484-332C-12	Sequence 12, Appl	809	14	0.8	822 4	US-09-662-250A-19	Sequence 19, Appl
C 737	14	0.8	619 6	5273901-4	Patent No. 5273901	810	14	0.8	827 4	US-09-662-250A-18	Sequence 18, Appl
C 738	14	0.8	619 6	5482709-3	Patent No. 5482709	811	14	0.8	828 4	US-09-404-879A-220	Sequence 220, App
C 739	14	0.8	624 4	US-08-961-527-324	Sequence 324, App	812	14	0.8	830 4	US-08-913-362-1	Sequence 1, Appl
C 740	14	0.8	628 1	US-08-434-255-18	Sequence 18, Appl	813	14	0.8	830 4	US-09-221-017B-1087	Sequence 1087, Ap
C 741	14	0.8	628 1	US-08-459-967-18	Sequence 18, Appl	814	14	0.8	833 2	US-08-790-137-2	Sequence 2, Appl
C 742	14	0.8	628 1	US-08-460-327-18	Sequence 18, Appl	815	14	0.8	833 2	US-08-913-362-5	Sequence 5, Appl
C 743	14	0.8	628 1	US-08-459-871-18	Sequence 18, Appl	816	14	0.8	833 2	US-08-913-362-5	Sequence 5, Appl
C 744	14	0.8	628 4	US-09-328-111-834	Sequence 834, App	817	14	0.8	834 4	US-09-187-789-6	Sequence 6, Appl
C 745	14	0.8	628 4	US-09-328-111-835	Sequence 328, App	818	14	0.8	834 4	US-09-187-789-6	Sequence 6, Appl
C 746	14	0.8	628 4	US-09-328-111-835	Sequence 328, App	819	14	0.8	834 4	US-09-187-789-6	Sequence 6, Appl
C 747	14	0.8	642 4	US-09-404-879A-325	Sequence 325, App	820	14	0.8	834 4	US-09-187-789-6	Sequence 6, Appl
C 748	14	0.8	646 4	US-08-961-527-328	Sequence 328, App	821	14	0.8	834 4	US-09-187-789-6	Sequence 6, Appl
C 749	14	0.8	646 4	US-09-443-906-5	Sequence 45, Appl	822	14	0.8	834 4	US-09-187-789-6	Sequence 6, Appl
C 750	14	0.8	649 1	US-08-463-202-23	Sequence 23, Appl	823	14	0.8	834 4	US-09-187-789-6	Sequence 6, Appl
C 751	14	0.8	650 1	US-08-463-202-23	Sequence 23, Appl	824	14	0.8	834 4	US-09-187-789-6	Sequence 6, Appl
C 752	14	0.8	650 2	US-08-484-434C-30	Sequence 30, Appl	825	14	0.8	834 4	US-09-187-789-6	Sequence 6, Appl
C 753	14	0.8	657 4	US-09-385-982-344	Sequence 344, App	826	14	0.8	834 4	US-09-187-789-6	Sequence 6, Appl
C 754	14	0.8	664 4	US-09-392-772-7	Sequence 7, Appl	827	14	0.8	834 4	US-09-187-789-6	Sequence 6, Appl
C 755	14	0.8	665 4	US-09-030-607-196	Sequence 196, App	828	14	0.8	834 4	US-09-187-789-6	Sequence 6, Appl
C 756	14	0.8	665 4	US-09-605-765-196	Sequence 196, App	829	14	0.8	834 4	US-09-187-789-6	Sequence 6, Appl
C 757	14	0.8	665 4	US-09-439-313-196	Sequence 196, App	830	14	0.8	834 4	US-09-187-789-6	Sequence 6, Appl



C 831	14	0.8	926	3	US-08-945-994-8	Sequence 8, Appli	C 904	14	0.8	1077	2	US-08-545-745-3	Sequence 3, Appli
C 832	14	0.8	942	4	US-09-470-512A-1	Sequence 1, Appli	C 905	14	0.8	1092	4	US-09-071-035-365	Sequence 365, App
C 833	14	0.8	942	4	US-09-134-001C-2348	Sequence 2348, Ap	906	14	0.8	1095	3	US-08-817-926-20	Sequence 20, Appl
C 834	14	0.8	949	3	US-08-714-918-12	Sequence 12, Appl	907	14	0.8	1095	4	US-09-203-893A-1	Sequence 1, Appli
C 835	14	0.8	949	4	US-09-265-315-12	Sequence 12, Appl	908	14	0.8	1095	4	US-09-203-893A-3	Sequence 3, Appli
C 836	14	0.8	949	4	US-09-265-315-12	Sequence 12, Appl	909	14	0.8	1095	4	US-09-203-893A-5	Sequence 5, Appli
C 837	14	0.8	949	4	US-09-266-417-12	Sequence 12, Appl	910	14	0.8	1095	4	US-09-203-893A-7	Sequence 7, Appli
C 838	14	0.8	950	2	US-08-501-200A-14	Sequence 14, Appl	911	14	0.8	1095	4	US-09-203-893A-9	Sequence 9, Appli
C 839	14	0.8	950	3	US-09-219-391-14	Sequence 14, Appl	912	14	0.8	1095	4	US-09-203-893A-11	Sequence 11, Appl
C 840	14	0.8	954	4	US-09-134-001C-875	Sequence 875, App	913	14	0.8	1095	4	US-09-203-893A-13	Sequence 13, Appl
C 841	14	0.8	954	4	US-09-134-001C-1156	Sequence 1156, Ap	914	14	0.8	1095	4	US-09-203-893A-15	Sequence 15, Appl
C 842	14	0.8	954	4	US-09-593-995-1	Sequence 1, Appli	915	14	0.8	1095	4	US-09-203-893A-17	Sequence 17, Appl
C 843	14	0.8	963	2	US-08-922-146-3	Sequence 3, Appli	916	14	0.8	1095	4	US-09-203-893A-19	Sequence 19, Appl
C 844	14	0.8	963	4	US-09-343-986-3	Sequence 3, Appli	917	14	0.8	1095	4	US-09-203-893A-21	Sequence 21, Appl
C 845	14	0.8	963	4	US-09-322-081-3	Sequence 3, Appli	918	14	0.8	1095	4	US-09-203-893A-23	Sequence 23, Appl
C 846	14	0.8	963	4	US-09-134-001C-1965	Sequence 1965, Ap	919	14	0.8	1095	4	US-09-203-893A-25	Sequence 25, Appl
C 847	14	0.8	963	4	US-09-221-017B-728	Sequence 728, App	920	14	0.8	1095	4	US-09-203-893A-27	Sequence 27, Appl
C 848	14	0.8	964	1	US-07-798-223A-1	Sequence 1, Appli	921	14	0.8	1095	4	US-09-203-893A-29	Sequence 29, Appl
C 849	14	0.8	966	2	US-08-922-146-1	Sequence 1, Appli	922	14	0.8	1095	4	US-09-203-893A-31	Sequence 31, Appl
C 850	14	0.8	966	4	US-09-343-986-1	Sequence 1, Appli	923	14	0.8	1095	4	US-09-134-001C-754	Sequence 754, App
C 851	14	0.8	966	4	US-09-322-081-1	Sequence 1, Appli	924	14	0.8	1101	4	US-09-134-001C-1720	Sequence 1720, Ap
C 852	14	0.8	966	4	US-09-540-014-27	Sequence 27, Appl	925	14	0.8	1111	1	US-08-202-186-15	Sequence 15, Appl
C 853	14	0.8	970	4	US-09-735-935-1	Sequence 1, Appli	926	14	0.8	1118	4	US-08-936-165A-156	Sequence 156, App
C 854	14	0.8	973	4	US-09-180-077-2	Sequence 2, Appli	927	14	0.8	1118	4	US-08-936-165A-35	Sequence 35, Appl
C 855	14	0.8	981	1	US-08-349-696-20	Sequence 20, Appl	928	14	0.8	1128	4	US-08-858-207A-85	Sequence 85, App
C 856	14	0.8	981	1	US-08-233-009-20	Sequence 20, Appl	929	14	0.8	1128	4	US-09-134-001C-696	Sequence 696, App
C 857	14	0.8	981	1	US-08-560-231-20	Sequence 20, Appl	930	14	0.8	1139	1	US-08-458-084-2	Sequence 2, Appli
C 858	14	0.8	981	4	US-09-080-704A-20	Sequence 20, Appl	931	14	0.8	1139	1	US-08-205-508-2	Sequence 2, Appli
C 859	14	0.8	983	4	US-09-564-805-216	Sequence 216, App	932	14	0.8	1139	1	US-08-278-630A-11	Sequence 11, Appl
C 860	14	0.8	988	4	US-09-282-146-1	Sequence 1, Appli	933	14	0.8	1139	2	US-08-482-148-7	Sequence 7, Appli
C 861	14	0.8	988	4	US-09-071-035-367	Sequence 367, App	934	14	0.8	1139	2	US-08-933-750C-71	Sequence 71, Appl
C 862	14	0.8	990	4	US-09-064-411A-8	Sequence 8, Appli	935	14	0.8	1139	3	US-09-234-613-71	Sequence 71, Appl
C 863	14	0.8	1000	4	US-09-222-938A-20	Sequence 20, Appl	936	14	0.8	1139	5	PCT-US95-02944-7	Sequence 7, Appli
C 864	14	0.8	1001	4	US-09-641-638-208	Sequence 208, App	937	14	0.8	1139	5	PCT-US95-02945-2	Sequence 2, Appli
C 865	14	0.8	1001	4	US-09-641-638-455	Sequence 455, App	938	14	0.8	1143	4	US-09-134-001C-993	Sequence 993, App
C 866	14	0.8	1002	1	US-08-463-090B-3	Sequence 3, Appli	939	14	0.8	1146	4	US-09-134-001C-906	Sequence 906, App
C 867	14	0.8	1002	4	US-09-328-111-159	Sequence 159, App	940	14	0.8	1152	4	US-09-032-215-24	Sequence 24, Appl
C 868	14	0.8	1002	4	US-09-465-353-1	Sequence 1, Appli	941	14	0.8	1152	4	US-09-032-215-25	Sequence 25, Appl
C 869	14	0.8	1003	1	US-07-800-364B-13	Sequence 13, Appl	942	14	0.8	1164	4	US-09-134-001C-411	Sequence 411, App
C 870	14	0.8	1003	2	US-07-989-847-11	Sequence 11, Appl	943	14	0.8	1166	1	US-08-599-252-95	Sequence 95, Appl
C 871	14	0.8	1003	4	US-08-469-411-11	Sequence 11, Appl	944	14	0.8	1166	4	US-08-944-483-7	Sequence 7, Appli
C 872	14	0.8	1008	4	US-09-354-138-139	Sequence 139, App	945	14	0.8	1166	5	PCT-US96-06352-95	Sequence 95, Appl
C 873	14	0.8	1013	6	524298-8	Patent No. 524298	946	14	0.8	1166	5	PCT-US96-06583-95	Sequence 95, Appl
C 874	14	0.8	1014	4	US-09-230-637-1	Sequence 1, Appli	947	14	0.8	1167	4	US-08-858-207A-69	Sequence 69, Appl
C 875	14	0.8	1023	2	US-08-757-653-175	Sequence 175, App	948	14	0.8	1167	4	US-09-134-001C-947	Sequence 947, App
C 876	14	0.8	1023	2	US-08-823-516-78	Sequence 78, Appl	949	14	0.8	1173	4	US-09-134-001C-281	Sequence 281, App
C 877	14	0.8	1023	3	US-08-759-038-114	Sequence 114, App	950	14	0.8	1174	2	US-08-733-410-9	Sequence 9, Appli
C 878	14	0.8	1023	3	US-08-758-314-114	Sequence 114, App	951	14	0.8	1177	2	US-08-892-770-3	Sequence 3, Appli
C 879	14	0.8	1023	4	US-09-724-517-3	Sequence 3, Appli	952	14	0.8	1183	4	US-09-227-357-20	Sequence 20, Appl
C 880	14	0.8	1023	4	US-09-641-807A-3	Sequence 3, Appli	953	14	0.8	1185	4	US-09-134-001C-1656	Sequence 1656, Ap
C 881	14	0.8	1024	4	US-09-723-096-3	Sequence 3, Appli	954	14	0.8	1185	4	US-08-944-483-8	Sequence 8, Appli
C 882	14	0.8	1031	1	US-08-064-411A-12	Sequence 12, Appl	955	14	0.8	1192	4	US-09-323-873A-1	Sequence 1, Appli
C 883	14	0.8	1031	1	US-08-181-271A-106	Sequence 106, App	956	14	0.8	1195	4	US-09-064-411A-27	Sequence 27, Appl
C 884	14	0.8	1031	1	US-08-449-315-106	Sequence 106, App	957	14	0.8	1197	4	US-09-063-743-2	Sequence 2, Appli
C 885	14	0.8	1031	1	US-08-444-803-106	Sequence 106, App	958	14	0.8	1211	4	US-09-590-540-2	Sequence 2, Appli
C 886	14	0.8	1031	1	US-08-449-043-106	Sequence 106, App	959	14	0.8	1211	4	PCT-US92-00282-8	Sequence 8, Appli
C 887	14	0.8	1031	1	US-08-456-265A-106	Sequence 106, App	960	14	0.8	1219	5	US-08-440-861-1	Sequence 1, Appli
C 888	14	0.8	1031	1	US-08-455-416-106	Sequence 106, App	961	14	0.8	1229	1	US-08-433-854-1	Sequence 1, Appli
C 889	14	0.8	1031	1	US-08-455-244-106	Sequence 106, App	962	14	0.8	1229	1	US-08-174-745A-1	Sequence 1, Appli
C 890	14	0.8	1031	1	US-08-454-876-106	Sequence 106, App	963	14	0.8	1229	1	US-08-195-947-1	Sequence 1, Appli
C 891	14	0.8	1031	2	US-08-457-364-106	Sequence 106, App	964	14	0.8	1229	2	US-08-433-885-1	Sequence 1, Appli
C 892	14	0.8	1031	2	US-08-456-262-106	Sequence 106, App	965	14	0.8	1229	2	US-08-433-908B-1	Sequence 1, Appli
C 893	14	0.8	1031	2	US-08-456-240-106	Sequence 106, App	966	14	0.8	1229	2	US-08-410-614-1	Sequence 1, Appli
C 894	14	0.8	1031	2	US-08-455-736-106	Sequence 106, App	967	14	0.8	1229	4	US-09-313-300-10	Sequence 10, Appl
C 895	14	0.8	1031	2	US-08-971-217-106	Sequence 106, App	968	14	0.8	1235	1	US-08-413-974-1	Sequence 1, Appli
C 896	14	0.8	1031	4	US-09-350-600-106	Sequence 106, App	969	14	0.8	1235	2	US-08-955-091-13	Sequence 13, Appl
C 897	14	0.8	1037	4	US-09-073-898-141	Sequence 141, App	970	14	0.8	1235	2	US-09-225-510-13	Sequence 13, Appl
C 898	14	0.8	1045	2	US-08-458-109-42	Sequence 42, Appl	971	14	0.8	1236	2	US-09-205-008-1	Sequence 1, Appli
C 899	14	0.8	1052	4	US-09-386-642-10	Sequence 10, Appl	972	14	0.8	1238	4	US-09-433-288-1	Sequence 1, Appli
C 900	14	0.8	1057	3	US-09-188-930-18	Sequence 18, Appl	973	14	0.8	1242	4	US-08-433-288-1	Sequence 1, Appli
C 901	14	0.8	1059	4	US-09-517-605-8	Sequence 8, Appli	974	14	0.8	1242	4	US-08-433-288-1	Sequence 1, Appli
C 902	14	0.8	1065	4	US-09-134-001C-2751	Sequence 2751, Ap	975	14	0.8	1242	4	US-08-433-288-1	Sequence 1, Appli
C 903	14	0.8	1067	2	US-08-854-170-1	Sequence 1, Appli	976	14	0.8	1242	4	US-08-433-288-1	Sequence 1, Appli

```
977 14 0 8 1242 4 US-08-174-739A-1 Sequence 1, Appl1
978 14 0 8 1242 4 US-08-434-256-1 Sequence 1, Appl1
979 14 0 8 1246 1 US-08-446-777-3 Sequence 3, Appl1
980 14 0 8 1247 4 US-09-399-913-15 Sequence 15, Appl1
981 14 0 8 1247 4 US-09-298-731-15 Sequence 15, Appl1
982 14 0 8 1257 4 US-09-064-411A-6 Sequence 6, Appl1
983 14 0 8 1257 4 US-09-134-001C-916 Sequence 916, App
984 14 0 8 1272 4 US-09-134-001C-2358 Sequence 2358, Ap
985 14 0 8 1276 4 US-09-064-411A-29 Sequence 29, Appl1
986 14 0 8 1278 4 US-09-347-803-3 Sequence 306, App
987 14 0 8 1279 1 US-08-146-010A-4 Sequence 4, Appl1
988 14 0 8 1279 1 US-08-674-168-9 Sequence 9, Appl1
989 14 0 8 1287 1 US-08-064-121-3 Sequence 3, Appl1
990 14 0 8 1287 1 US-08-478-015-3 Sequence 3, Appl1
991 14 0 8 1287 3 US-08-475-975-3 Sequence 3, Appl1
992 14 0 8 1287 3 US-09-084-889-3 Sequence 3, Appl1
993 14 0 8 1287 3 US-09-084-889-3 Sequence 3, Appl1
994 14 0 8 1295 1 US-08-245-295-6 Sequence 6, Appl1
995 14 0 8 1295 1 US-08-481-130-6 Sequence 6, Appl1
996 14 0 8 1295 1 US-08-656-984A-6 Sequence 6, Appl1
997 14 0 8 1295 1 US-08-485-604-6 Sequence 6, Appl1
998 14 0 8 1295 2 US-08-487-595-6 Sequence 6, Appl1
999 14 0 8 1295 3 US-08-863-790-25 Sequence 25, Appl1
1000 14 0 8 1295 3 US-08-296-749-25 Sequence 25, Appl1
```

## ALIGNMENTS

```
RESULT 1
US-08-720-484A-3/C
; Sequence 3, Application US/08720484A
; Patent No. 5990281
; GENERAL INFORMATION:
; APPLICANT: Desauvage, Frederic
; APPLICANT: Rosenthal, Arnon
; APPLICANT: Stone, Donna
; TITLE OF INVENTION: Vertebrate Smoothened Proteins
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/720,484A
; FILING DATE: 30-Sep-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1050
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1469
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2972 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-720-484A-3

Query Match 1.0%; Score 19; DB 2; Length 2972;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1280 TTGGCAATCATCTTGCTCT 1298
Db 1724 TTGGCAATCATCTTGCTCT 1706
```

```
RESULT 2
US-08-953-823A-3/C
; Sequence 3, Application US/08953823A
; Patent No. 616958
; GENERAL INFORMATION:
; APPLICANT: Desauvage, Frederic
; APPLICANT: Rosenthal, Arnon
; APPLICANT: Stone, Donna
; TITLE OF INVENTION: Vertebrate Smoothened Proteins
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/953,823A
; FILING DATE: 30-Sep-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/027070
; FILING DATE: 30-Sep-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1050R1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2972 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-953-823A-3

Query Match 1.0%; Score 19; DB 3; Length 2972;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1280 TTGGCAATCATCTTGCTCT 1298
Db 1724 TTGGCAATCATCTTGCTCT 1706

RESULT 3
US-09-398-239-3/C
; Sequence 3, Application US/09398239
; Patent No. 6407216
; GENERAL INFORMATION:
; APPLICANT: Desauvage, Frederic
; APPLICANT: Rosenthal, Arnon
; APPLICANT: Stone, Donna
; TITLE OF INVENTION: Vertebrate Smoothened Proteins
; FILE REFERENCE: P1050D1
; CURRENT APPLICATION NUMBER: US/09/398,239
; CURRENT FILING DATE: 1999-09-20
; EARLIER APPLICATION NUMBER: US 08/720,484
; EARLIER FILING DATE: 1996-09-30
```

```
; NUMBER OF SEQ ID NOS: 5
; SEQ ID NO 3
; LENGTH: 2972
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-398-239-3

Query Match      1.0%; Score 19; DB 4; Length 2972;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1280 TTGGCAATCATCTGCTCT 1298
|||||
Db 1724 TTGGCAATCATCTGCTCT 1706
|||||

RESULT 4
US-09-134-001C-182
; Sequence 182, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 182
; LENGTH: 378
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-182

Query Match      1.0%; Score 18; DB 4; Length 378;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 266 AAAAAACAATGTACAA 283
|||||
Db 211 AAAAAACAATGTACAA 228
|||||

RESULT 5
US-08-343-101A-8
; Sequence 8, Application US/08343101A
; Patent No. 5830759
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
; TITLE OF INVENTION: Virus Sequences And Uses Thereof
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,101A
; FILING DATE:
; CLASSIFICATION: 514
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq., John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0526
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1315 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
US-08-343-101A-8

Query Match      1.0%; Score 18; DB 2; Length 1315;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 396 TCTGGCAGCTCTACAGT 413
|||||
Db 526 TCTGGCAGCTCTACAGT 543
|||||

RESULT 6
US-09-183-688-8
; Sequence 8, Application US/09183688
; Patent No. 6093550
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
; TITLE OF INVENTION: Virus Sequences And Uses Thereof
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/183,688
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/343,101
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq., John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0526
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1315 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
US-09-183-688-8
```

Query Match 1.0%; Score 18; DB 3; Length 1315;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 TCTGGGAGCTCTACAGT 413  
DB 526 TCTGGGAGCTCTACAGT 543

## RESULT 7

US-09-325-932A-102  
Sequence 102, Application US/09325932A  
Patent No. 6451604  
GENERAL INFORMATION:  
APPLICANT: Flinn, Barry  
APPLICANT: Lasham, Annette  
TITLE OF INVENTION: Compositions affecting programmed cell  
TITLE OF INVENTION: death and their use in the modification of forestry plant devel  
FILE REFERENCE: 1022  
CURRENT APPLICATION NUMBER: US/09/325,932A  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 206  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 102  
LENGTH: 2167  
TYPE: DNA  
ORGANISM: Pinus radiata  
US-09-325-932A-102

Query Match 1.0%; Score 18; DB 4; Length 2167;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 655 GTTGGCTTGACATTGAT 672  
DB 1610 GTTGGCTTGACATTGAT 627

## RESULT 8

US-08-420-235B-20  
Sequence 20, Application US/08420235B  
Patent No. 5801042

GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS  
TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/420,235B  
FILING DATE:  
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 45185-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 20:

## SEQUENCE CHARACTERISTICS:

LENGTH: 2259 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: N  
ANTI-SENSE: N  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2259  
OTHER INFORMATION:

US-08-420-235B-20

Query Match 1.0%; Score 18; DB 1; Length 2259;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 TCTGGGAGCTCTACAGT 413  
DB 1194 TCTGGGAGCTCTACAGT 1211

## RESULT 9

US-08-793-624-20  
Sequence 20, Application US/08793624C  
Patent No. 6150093  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And  
TITLE OF INVENTION: Uses Thereof  
FILE REFERENCE: 45185-C-PCT-US/JPW  
CURRENT APPLICATION NUMBER: US/08/793,624C  
CURRENT FILING DATE: 1997-02-18  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 20  
LENGTH: 2259  
TYPE: DNA  
ORGANISM: Kaposi's sarcoma-associated herpesvirus  
US-08-793-624-20

Query Match 1.0%; Score 18; DB 3; Length 2259;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 TCTGGGAGCTCTACAGT 413  
DB 1194 TCTGGGAGCTCTACAGT 1211

## RESULT 10

PCT-US95-10194-20  
Sequence 20, Application PC/TUS9510194  
GENERAL INFORMATION:

APPLICANT: The Trustees of Columbia University in the City of New York  
APPLICANT: City  
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS  
TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER:  PCT/US95/10194
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME:  White, John P.
REGISTRATION NUMBER:  28,678
REFERENCE/DOCKET NUMBER:  45185-C-PCT/JPW/MSC
TELECOMMUNICATION INFORMATION:
TELEPHONE:  (212) 278-0400
TELEFAX:  (212) 391-0525
INFORMATION FOR SEQ ID NO:  20:
SEQUENCE CHARACTERISTICS:
LENGTH:  2259 base pairs
TYPE:  nucleic acid
STRANDEDNESS:  single
TOPOLOGY:  linear
MOLECULE TYPE:  DNA (genomic)
HYPOTHETICAL:  N
ANTI-SENSE:  N
FEATURE:
NAME/KEY:  CDS
LOCATION:  1..2259
OTHER INFORMATION:
PCT-US95-10194-20

```

Query Match 1.0%; Score 18; DB 4; Length 4796;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 252 CTTCTGATCTTGGGAAA 269  
Db 4717 CTTCTGATCTTGGGAAA 4734

RESULT 13  
US-08-420-235B-1/c  
Sequence 1, Application US/08420235B  
Patent No. 5801042  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS  
TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/420,235B  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 45185-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20710 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: N  
ANTI-SENSE: N  
US-08-420-235B-1

Query Match 1.0%; Score 18; DB 1; Length 20710;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 396 TCTGGGACGCTCTACAGT 413  
Db 6565 TCTGGGACGCTCTACAGT 5548  
RESULT 14  
US-08-793-624-1/c  
Sequence 1, Application US/08793624C  
Patent No. 6150093  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And  
TITLE OF INVENTION: Uses Thereof  
FILE REFERENCE: 45185-C-PCT-US/JPW  
CURRENT APPLICATION NUMBER: US/08/793,624C

CURRENT FILING DATE: 1997-02-18  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 1  
LENGTH: 20710  
TYPE: DNA  
ORGANISM: Kaposi's sarcoma-associated herpesvirus  
US-08-793-624-1

Query Match 1.0%; Score 18; DB 3; Length 20710;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 TCTGGGACGCTCTACAGT 413  
Db 6565 TCTGGGACGCTCTACAGT 6548

RESULT 15  
PCT-US95-10194-1/c  
Sequence 1, Application PC/TUS9510194  
GENERAL INFORMATION:  
APPLICANT: The Trustees of Columbia University in the City of New York  
APPLICANT: City  
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS  
TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/10194  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MSC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20710 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: N  
ANTI-SENSE: N  
PCT-US95-10194-1

Query Match 1.0%; Score 18; DB 5; Length 20710;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 TCTGGGACGCTCTACAGT 413  
Db 6565 TCTGGGACGCTCTACAGT 6548

RESULT 16  
US-08-961-527-83  
Sequence 83, Application US/08961527

```
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28473 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-83

Query Match 1.0%; Score 18; DB 4; Length 28473;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 GAAAAATCATGCCAGTGT 150
Db 17357 GAAAAATCATGCCAGTGT 17374

RESULT 17
US-08-770-379-18/c
; Sequence 18, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-757-669A-18
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-770-379-18

Query Match 1.0%; Score 18; DB 2; Length 35100;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 396 TCTGGGCAGCTCTACAGT 413
Db 6485 TCTGGGCAGCTCTACAGT 6468

RESULT 18
US-08-757-669A-18/c
; Sequence 18, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-757-669A-18
```

Query Match 1.0%; Score 18; DB 4; Length 35100;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 TCTGGGAGGCTCTACAGT 413  
DB 6485 TCTGGGAGGCTCTACAGT 6468

RESULT 19  
US-09-230-371A-18/c  
; Sequence 18, Application US/09230371A  
; Patent No. 6348586

; GENERAL INFORMATION:  
; APPLICANT: Chang, Yuan  
; APPLICANT: Bohenzky, Roy J.  
; APPLICANT: Russo, James S  
; APPLICANT: Edelman, Isidore S  
; APPLICANT: Moore, Patrick S  
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND  
; FILE REFERENCE: 45185-G-PCT-US  
; CURRENT APPLICATION NUMBER: US/09/230,371A  
; CURRENT FILING DATE: 1999-11-17  
; PRIOR APPLICATION NUMBER: PCT/US97/13346  
; PRIOR FILING DATE: 1997-07-22  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 35100  
; TYPE: DNA  
; ORGANISM: Kaposi's sarcoma-associated herpesvirus  
US-09-230-371A-18

Query Match 1.0%; Score 18; DB 4; Length 35100;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 TCTGGGAGGCTCTACAGT 413  
DB 6485 TCTGGGAGGCTCTACAGT 6468

RESULT 20  
US-09-749-588-3/c  
; Sequence 3, Application US/09749588  
; Patent No. 6423521

; GENERAL INFORMATION:  
; APPLICANT: CHANDRAMOULISWARAN, Ishwar et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CU001068  
; CURRENT APPLICATION NUMBER: US/09/749,588  
; CURRENT FILING DATE: 2000-12-28  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 36159  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(36159)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-749-588-3

Query Match 1.0%; Score 18; DB 4; Length 36159;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 829 TGTTTCTGGAAGTCTTT 846

DB 27845 TGTTTCTGGAAGTCTTT 27828

RESULT 21  
US-09-090-793-12  
; Sequence 12, Application US/09090793  
; Patent No. 6140466

; GENERAL INFORMATION:  
; APPLICANT: Calgene, LLC  
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression  
; TITLE OF INVENTION: of polyketide-like synthesis genes in plants  
; FILE REFERENCE: CGN 131.01US  
; CURRENT APPLICATION NUMBER: US/09/090,793  
; CURRENT FILING DATE: 1998-06-04  
; EARLIER APPLICATION NUMBER: 60/048,650  
; EARLIER FILING DATE: 1997-06-04  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 40138  
; TYPE: DNA  
; ORGANISM: Vibrio marinus  
US-09-090-793-12

Query Match 1.0%; Score 18; DB 3; Length 40138;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 TTTTGATGAGCGTATATT 341  
DB 551 TTTTGATGAGCGTATATT 568

RESULT 22  
US-08-756-771-2/c  
; Sequence 2, Application US/08756771  
; Patent No. 5817497

; GENERAL INFORMATION:  
; APPLICANT: Goll, Surya K.  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: A NOVEL GLUTATHIONE S-TRANSFERASE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/756,771  
; FILING DATE: Herewith  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0162 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 800 base pairs



```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
US-08-756-771-2

Query Match      0.9%; Score 17; DB 1; Length 800;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 674 TTAGGAATTTCTCTTC 690
Db 518 TTAGGAATTTCTCTTC 502

RESULT 23
US-09-096-571-2/c
; Sequence 2, Application US/09096571
; Patent No. 5976528
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL GLUTATHIONE S-TRANSFERASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/096,571
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/096,571
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0162 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0162 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
US-09-096-571-2

Query Match      0.9%; Score 17; DB 2; Length 800;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 674 TTAGGAATTTCTCTTC 690
Db 518 TTAGGAATTTCTCTTC 502
```

```
RESULT 24
US-09-309-320-2/c
; Sequence 2, Application US/09309320
; Patent No. 6248325
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL GLUTATHIONE S-TRANSFERASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/309,320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/096,571
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0162 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
US-09-309-320-2

Query Match      0.9%; Score 17; DB 4; Length 800;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 674 TTAGGAATTTCTCTTC 690
Db 518 TTAGGAATTTCTCTTC 502

RESULT 25
US-08-617-8608-21
; Sequence 21, Application US/086178608
; Patent No. 6133506
; GENERAL INFORMATION:
; APPLICANT: Typer, R., Bautor, J., Bothmann, H., Filsak, E.,
; APPLICANT: Hrvick-Grandpierre, C., Klein, B., Martini, N.,
; APPLICANT: Mller, A., Schulte, W., Voetz, M., Walek, J.,
; APPLICANT: Schell, J.
; TITLE OF INVENTION: Promoters
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
; STREET: 1140 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
```

ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (PFO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/617, 8608  
FILING DATE: 01-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP94/02950  
FILING DATE: 05-SEP-1994  
APPLICATION NUMBER: DE P4329951.2  
FILING DATE: 04-SEP-1993  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 926 base pairs  
TYPE: Nucleic acid  
STRANDEDNESS: Double stranded  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Cuphea lanceolata  
IMMEDIATE SOURCE:  
LIBRARY: genomic Lambda FIX II  
CLONE: C1KRg3  
FEATURE:  
NAME/KEY: TATA-Signal  
LOCATION: 827..838  
FEATURE:  
NAME/KEY: Transcription start  
LOCATION: 864  
FEATURE:  
NAME/KEY: Startcodon  
LOCATION: 916..918  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 916..926  
US-08-617-8608-21

Query Match 0.9%; Score 17; DB 3; Length 926;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 CTTTGTCCCGAGGT 232  
DB 459 CTTTGTCCCGAGGT 475

RESULT 26  
US-09-045-583-6  
Sequence 6, Application US/09045583  
Patent No. 6287805  
GENERAL INFORMATION:  
APPLICANT: Graham, Gerard J. et al.  
TITLE OF INVENTION: No. 6287805el Molecules of the G Protein-Coupled  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHYE & COCKFIELD, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/045,583

FILING DATE: 20-MAR-98  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: MNI-044  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1137 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1134  
US-09-045-583-6

Query Match 0.9%; Score 17; DB 4; Length 1137;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 720 TTTTCTCATGCTGTGG 736  
DB 770 TTTTCTCATGCTGTGG 786

RESULT 27  
US-09-534-185-6  
Sequence 6, Application US/09534185  
Patent No. 6403767  
GENERAL INFORMATION:  
APPLICANT: Graham, Gerard J. et al.  
TITLE OF INVENTION: No. 6403767el Molecules of the G Protein-Coupled  
Heptahelical Receptor Superfamily and Uses  
Therefor  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHYE & COCKFIELD, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/534,185  
FILING DATE: 24-Mar-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/045,583  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: MNI-044  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1137 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1134  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-534-185-6

Query Match 0.9%; Score 17; DB 4; Length 1137;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 720 TTTTCTTCATGCTGTGG 736  
|||||  
Db 770 TTTTCTTCATGCTGTGG 786

RESULT 28  
US-09-045-583-4  
; Sequence 4, Application US/09045583  
; Patent No. 6287805  
; GENERAL INFORMATION:  
; APPLICANT: Graham, Gerard J. et al.  
; TITLE OF INVENTION: No. 6287805el Molecules of the G Protein-Coupled  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD, LLP  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/045,583  
FILING DATE: 20-MAR-98  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: MNI-044  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1664 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:

NAME/KEY: CDS  
LOCATION: 193..1326  
US-09-045-583-4  
Query Match 0.9%; Score 17; DB 4; Length 1664;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 720 TTTTCTTCATGCTGTGG 736  
|||||  
Db 962 TTTTCTTCATGCTGTGG 978

RESULT 29  
US-09-534-185-4  
; Sequence 4, Application US/09534185  
; Patent No. 6403767  
; GENERAL INFORMATION:  
; APPLICANT: Graham, Gerard J. et al.  
; TITLE OF INVENTION: No. 6403767el Molecules of the G Protein-Coupled  
; Heptahelical Receptor Superfamily and Uses  
; Therefor

NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/534,185  
FILING DATE: 24-Mar-2000

CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/045,583  
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: MNI-044  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1664 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:

NAME/KEY: CDS  
LOCATION: 193..1326  
US-09-534-185-4

Query Match 0.9%; Score 17; DB 4; Length 1664;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 720 TTTTCTTCATGCTGTGG 736  
|||||  
Db 962 TTTTCTTCATGCTGTGG 978

RESULT 30  
US-09-134-001C-79  
; Sequence 79, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lyvin Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 79  
LENGTH: 1692  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-79

Query Match 0.9%; Score 17; DB 4; Length 1692;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1210 TTGCTCAAGGTAGCA 1226  
DB 799 TTGCTCAAGGTAGCA 815

## RESULT 31

US-08-936-165A-116/c  
Sequence 116, Application US/08936165A  
Patent No. 6348582

GENERAL INFORMATION:  
APPLICANT: Black, Michael  
APPLICANT: Burnham, Martin  
APPLICANT: Hodgson, John  
APPLICANT: Knowles, David  
APPLICANT: Lometto, Michael  
APPLICANT: Nicholas, Richard  
APPLICANT: Pratt, Julie  
APPLICANT: Reichard, Richard  
APPLICANT: Rosenberg, Martin  
APPLICANT: Ward, Judith  
TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,  
TITLE OF INVENTION: Polypeptides and Their Uses  
NUMBER OF SEQUENCES: 534  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/936,165A  
FILING DATE: 24-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027,032  
FILING DATE: 24-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Giumi, Edward R.  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 116:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1789 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-08-936-165A-116

Query Match 0.9%; Score 17; DB 4; Length 1789;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 CCTCAGATTTAAATCT 205  
DB 94 CCTCAGATTTAAATCT 78

## RESULT 32

US-09-484-970B-165  
Sequence 165, Application US/09484970B  
Patent No. 6426186

GENERAL INFORMATION:  
APPLICANT: Jones, Karen A.  
APPLICANT: Volkmuth, Wayne  
APPLICANT: Walker, Michael G.  
TITLE OF INVENTION: BONE REMODELING GENES  
FILE REFERENCE: PB-0014 US  
CURRENT APPLICATION NUMBER: US/09/484,970B  
CURRENT FILING DATE: 2000-01-18  
NUMBER OF SEQ ID NOS: 172  
SOFTWARE: PERL Program  
SEQ ID NO 165  
LENGTH: 214  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. 6426186 199021.1  
NAME/KEY: unsure  
LOCATION: 1907, 1909, 2009, 2014  
OTHER INFORMATION: a, t, c, g, or other  
US-09-484-970B-165

Query Match 0.9%; Score 17; DB 4; Length 214;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 TTAATAATCTGAGAA 553  
DB 1449 TTAATAATCTGAGAA 1465

## RESULT 33

US-09-066-047-17/c  
Sequence 17, Application US/09066047A  
Patent No. 6306394

GENERAL INFORMATION:  
APPLICANT: MURPHY, Cheryl  
STOREY, James  
BEUTZ, Gerald A.  
TITLE OF INVENTION: COUGH-LIN, Richard T.  
TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF  
USE OF GRANULOCYTIC ERLHICHA  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HALE AND DORR LLP  
STREET: 60 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: United States  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,047A  
FILING DATE: 24-Apr-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/044,869  
FILING DATE: 25-APR-1997  
ATTORNEY/AGENT INFORMATION:

NAME: Superko, Colleen  
REGISTRATION NUMBER: 39,850  
REFERENCE/DOCKET NUMBER: 106,941,156  
TELEPHONE: (617) 526-6000  
TELEFAX: (617) 526-5000  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2900 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 17:  
US-09-066-047-17

Query Match 0.9%; Score 17; DB 4; Length 2900;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 192 CAGATTAAATCTGGA 208  
Db 1674 CAGATTAAATCTGGA 1658

RESULT 34  
US-08-720-484A-1/c  
; Sequence 1, Application US/08720484A  
; Patent No. 5990281  
; GENERAL INFORMATION:  
; APPLICANT: Desauvage, Frederic  
; APPLICANT: Rosenthal, Arnon  
; APPLICANT: Stone, Donna  
; TITLE OF INVENTION: Vertebrate Smoothened Proteins  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/720,484A  
; FILING DATE: 30-Sep-1996  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1050  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3854 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
US-08-720-484A-1

Query Match 0.9%; Score 17; DB 2; Length 3854;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1280 TTGGCAATCATCTTGCT 1296

Db 2173 TTGGCAATCATCTTGCT 2157  
RESULT 35  
US-08-953-823A-1/c  
; Sequence 1, Application US/08953823A  
; Patent No. 6136958  
; GENERAL INFORMATION:  
; APPLICANT: Desauvage, Frederic  
; APPLICANT: Rosenthal, Arnon  
; APPLICANT: Stone, Donna  
; TITLE OF INVENTION: Vertebrate Smoothened Proteins  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/953,823A  
; FILING DATE: 30-Sep-1997  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/027070  
; FILING DATE: 30-Sep-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1050R1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3854 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
US-08-953-823A-1

Query Match 0.9%; Score 17; DB 3; Length 3854;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1280 TTGGCAATCATCTTGCT 1296  
Db 2173 TTGGCAATCATCTTGCT 2157

RESULT 36  
US-09-398-239-1/c  
; Sequence 1, Application US/09398239  
; Patent No. 6407216  
; GENERAL INFORMATION:  
; APPLICANT: Desauvage, Frederic  
; APPLICANT: Rosenthal, Arnon  
; APPLICANT: Stone, Donna  
; TITLE OF INVENTION: Vertebrate Smoothened Proteins  
; FILE REFERENCE: P1050D1  
; CURRENT APPLICATION NUMBER: US/09/398,239  
; CURRENT FILING DATE: 1999-09-20  
; EARLIER APPLICATION NUMBER: US 08/720,484  
; EARLIER FILING DATE: 1996-09-30  
; NUMBER OF SEQ ID NOS: 5  
; SEQ ID NO 1



```
; CURRENT APPLICATION NUMBER: US/09/295,028
; CURRENT FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 7091
; TYPE: DNA
; ORGANISM: Ehrlichia sp.
US-09-295-028-46

Query Match          0.9%; Score 17; DB 4; Length 7091;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1781 CGGTAGCAATATAAA 1797
   |||||
Db 4451 CGGTAGCAATATAAA 4435

RESULT 40
US-09-106-582-46/c
; Sequence 46, Application US/09106582
; Patent No. 6306402
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; THERAPY
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,582
; FILING DATE: 29-JUN-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.439C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7091 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; MOLECULE TYPE: DNA (genomic)
US-09-106-582-46

Query Match          0.9%; Score 17; DB 4; Length 7091;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1781 CGGTAGCAATATAAA 1797
   |||||
Db 4451 CGGTAGCAATATAAA 4435

RESULT 41
US-08-961-527-189/c
; Sequence 189, Application US/08961527
```

```
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 189:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7174 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-189

Query Match          0.9%; Score 17; DB 4; Length 7174;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 531 AAAGCTTTAAAAATCT 547
   |||||
Db 2311 AAAGCTTTAAAAATCT 2295

RESULT 42
US-08-961-527-167
; Sequence 167, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
```

APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Andrews  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 167:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9711 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-167

Query Match 0.9%; Score 17; DB 4; Length 9711;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1762 ACCGTAACCTATCT 1778  
DB 7007 ACCGTAACCTATCT 7023

RESULT 43  
US-08-038-768A-4/c  
Sequence 4, Application US/08038768A  
Patent No. 5569828  
GENERAL INFORMATION:  
APPLICANT: McMillen, Michael D.; Roth, Bradley A.; Townsend,  
APPLICANT: Rod  
TITLE OF INVENTION: MAIZE CHLOROTIC DWARF VIRUS RESISTANCE  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pioneer Hi-Bred International, Inc.  
STREET: 700 Capital Square, 400 Locust  
CITY: Des Moines  
STATE: Iowa  
COUNTRY: United States  
ZIP: 50309  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS, Microsoft Windows  
SOFTWARE: Microsoft Windows No. 5569828eepad  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/038.768A  
FILING DATE: 19930324  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Roth, Michael J.  
REGISTRATION NUMBER: 29,342  
REFERENCE/DOCKET NUMBER: 0235 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (515) 245-3594  
TELEFAX: (515) 245-3634  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11785 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: viral RNA  
HYPOTHETICAL: No  
US-08-038-768A-4  
Query Match 0.9%; Score 17; DB 1; Length 11785;

Best Local Similarity 100.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1736 CGTAAGAAAATCGCTC 1752  
DB 965 CGTAAGAAAATCGCTC 949

RESULT 44  
US-08-416-603-3/c  
Sequence 3, Application US/08416603  
Patent No. 5866780  
GENERAL INFORMATION:  
APPLICANT: Law, Marcus  
APPLICANT: Hebata, Uedare  
APPLICANT: Reddick, Bradford B.  
TITLE OF INVENTION: Maize Chlorotic Dwarf Virus Genome and  
TITLE OF INVENTION: Uses Therefor  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/416,603  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lloyd, Jeffrey  
REGISTRATION NUMBER: 35,589  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11785 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1..11785  
OTHER INFORMATION: /note="CDNA sequence of MCDV-T  
OTHER INFORMATION: genome as disclosed in WO 94/21796"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 456..10826  
US-08-416-603-3

Query Match 0.9%; Score 17; DB 2; Length 11785;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1736 CGTAAGAAAATCGCTC 1752  
DB 965 CGTAAGAAAATCGCTC 949

RESULT 45  
US-08-687-080-105/c  
Sequence 105, Application US/08687080  
Patent No. 5965427  
GENERAL INFORMATION:  
APPLICANT: Gregory Dolganov







GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 08:59:21 ; Search time 3371 Seconds  
(without alignments)  
16092.444 Million cell updates/sec

Title: US-09-868-987-1  
Perfect score: 1864  
Sequence: 1 atggacttcgcattgttc.....atgacaaattccagataatgc 1864

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.Other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.ey.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.Other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1453	78.0	14134	1	AE001641 Chlamydia
C 2	1453	78.0	300550	1	AP002547 Chlamydia
C 3	1421	76.2	12198	1	AE002179 Chlamydia
C 4	92	4.9	11938	1	AE002180 Chlamydia
C 5	22	1.2	3183	5	AF019977 Gallus ga
C 6	22	1.2	151032	9	AC105201 Homo sapi
C 7	22	1.2	159198	2	AC124363 Mus muscu
C 8	22	1.2	172579	9	AC036125 Homo sapi
C 9	22	1.2	209636	2	AC121084 Mus muscu
C 10	22	1.2	247074	2	AC121990 Mus muscu
C 11	22	1.2	293272	2	AC120685 Rattus no
C 12	21	1.1	15389	1	AF026147 Bacillus
C 13	21	1.1	41830	3	CBRG44E08
C 14	21	1.1	90157	9	AL355552 Human DNA
C 15	21	1.1	94960	9	AC000066 Homo sapi
C 16	21	1.1	100170	9	AL161743 Human DNA
C 17	21	1.1	110000	2	AC122235_2
C 18	21	1.1	114741	2	AC124954
C 19	21	1.1	123091	2	AC129991
C 20	21	1.1	124028	2	AC137017 Rattus no
C 21	21	1.1	133054	2	AL445200 Homo sapi
C 22	21	1.1	145661	2	AC114333 Canis fam
C 23	21	1.1	147803	9	AL356867 Human DNA
C 24	21	1.1	157326	2	AC011024 Homo sapi
C 25	21	1.1	168632	2	AC106409 Rattus no
C 26	21	1.1	170008	9	HS010770 Homo sapi
C 27	21	1.1	175809	2	AC096466 Rattus no
C 28	21	1.1	207730	1	BSUB0011
C 29	21	1.1	207890	2	AL732291 Mus muscu
C 30	21	1.1	219622	9	AC009630 Homo sapi
C 31	21	1.1	257967	2	AL365337 Mus muscu
C 32	20	1.1	1472	3	AF322045 Malawimon
C 33	20	1.1	2895	1	CAU38234
C 34	20	1.1	3644	3	IGU48714
C 35	20	1.1	7370	5	AF043383 Pleuronec
C 36	20	1.1	9989	1	AE001318 Chlamydia
C 37	20	1.1	10742	1	AE007683 Clostridi
C 38	20	1.1	10836	1	AE014001 Yersinia
C 39	20	1.1	12442	1	AE002341 Chlamydia
C 40	20	1.1	42632	3	CEP52H3
C 41	20	1.1	72371	2	AC119990
C 42	20	1.1	73256	2	AC069119 Homo sapi
C 43	20	1.1	74105	2	AC087648 Homo sapi
C 44	20	1.1	84574	9	AL713969 Human DNA
C 45	20	1.1	101730	2	AC111865 Rattus no
C 46	20	1.1	102086	9	AL512447 Human DNA
C 47	20	1.1	102846	9	AC002483
C 48	20	1.1	111357	10	AC110422
C 49	20	1.1	113425	2	AP004256
C 50	20	1.1	127334	9	AC109579 Homo sapi
C 51	20	1.1	133072	9	HS1097P24
C 52	20	1.1	134703	2	AC124259 Homo sapi
C 53	20	1.1	135373	2	AP004695
C 54	20	1.1	135378	9	AL390060 Human DNA
C 55	20	1.1	138824	9	AC008689 Homo sapi
C 56	20	1.1	141267	2	AC013694
C 57	20	1.1	142100	2	AC094067
C 58	20	1.1	142600	9	AC005371 Homo sapi
C 59	20	1.1	143024	2	AC098435
C 60	20	1.1	146528	2	AC106923 Rattus no
C 61	20	1.1	147009	2	AC084179 Homo sapi
C 62	20	1.1	150204	8	OSJN00136
C 63	20	1.1	154761	2	AP005312
C 64	20	1.1	159791	9	AC007179 Homo sapi
C 65	20	1.1	160799	2	AC021601 Homo sapi

66	20	1.1	160813	2	AC099502	AC099502 Homo sapi	c 139	19	1.0	84129	8	AB013392	AB013392 Arabidops
67	20	1.1	162460	2	AC120295	AC120295 Rattus no	c 140	19	1.0	85791	8	AB016888	AB016888 Arabidops
68	20	1.1	163198	9	AC079175	AC079175 Homo sapi	c 141	19	1.0	87636	9	AL139039	AL139039 Human DNA
69	20	1.1	165565	9	AC073651	AC073651 Homo sapi	c 142	19	1.0	90352	2	AC099278	AC099278 Rattus no
70	20	1.1	165565	9	AL445212	AL445212 Human DNA	c 143	19	1.0	91350	9	HSJ586J1	HSJ586J1 Human DNA
71	20	1.1	166566	2	AC109133	AC109133 Homo sapi	c 144	19	1.0	92948	8	ATP00657	ATP00657 Arabidops
72	20	1.1	169228	2	AC096446	AC096446 Rattus no	c 145	19	1.0	93163	9	HSAC000121	HSAC000121 Human BAC
73	20	1.1	176253	2	AP001078	AP001078 Homo sapi	c 146	19	1.0	93756	10	AL512588	AL512588 Mouse DNA
74	20	1.1	176610	2	AC099371	AC099371 Rattus no	c 147	19	1.0	96334	2	AC021343	AC021343 Homo sapi
75	20	1.1	179202	2	AC025935	AC025935 Homo sapi	c 148	19	1.0	96334	2	AC021343	AC021343 Homo sapi
76	20	1.1	180445	2	AC019171	AC019171 Homo sapi	c 149	19	1.0	97255	9	AL590482	AL590482 Human DNA
77	20	1.1	181503	2	AC093381	AC093381 Rattus no	c 150	19	1.0	98939	8	AC025435	AC025435 Homo sapi
78	20	1.1	187635	2	AC119381	AC119381 Rattus no	c 151	19	1.0	103568	8	CNS078G7	CNS078G7 Homo centi
79	20	1.1	190556	2	AC020733	AC020733 Homo sapi	c 152	19	1.0	105742	2	AC010438	AC010438 Homo sapi
80	20	1.1	190948	2	AC093138	AC093138 Pan trogl	c 153	19	1.0	106295	8	AC115378	AC115378 Rattus no
81	20	1.1	192524	2	AL732504	AL732504 Mus muscu	c 154	19	1.0	106355	8	AC005309	AC005309 Arabidops
82	20	1.1	193278	2	AL773590	AL773590 Mus muscu	c 155	19	1.0	106902	9	AC080011	AC080011 Homo sapi
83	20	1.1	194297	2	AC044810	AC044810 Homo sapi	c 156	19	1.0	111185	2	AC126010	AC126010 Medicago
84	20	1.1	196374	2	CNS01D07	AL133168 Homo sapi	c 157	19	1.0	118244	10	AL645759	AL645759 Mouse DNA
85	20	1.1	198134	2	AC024117	AC024117 Homo sapi	c 158	19	1.0	119509	2	CNS08C9L	AL732646 Oryza sat
86	20	1.1	205639	9	AP001793	AP001793 Homo sapi	c 159	19	1.0	128885	9	AP004318	AP004318 Homo sapi
87	20	1.1	206160	2	AC113011	AC113011 Mus muscu	c 160	19	1.0	130059	8	AC078840	AC078840 Oryza sat
88	20	1.1	206266	2	AC109220	AC109220 Mus muscu	c 161	19	1.0	132075	2	AC112539	AC112539 Rattus no
89	20	1.1	210734	2	AC015676	AC015676 Homo sapi	c 162	19	1.0	133457	9	AL161624	AL161624 Human DNA
90	20	1.1	212055	2	AP000899	AP000899 Homo sapi	c 163	19	1.0	133829	2	AC026242	AC026242 Homo sapi
91	20	1.1	226431	2	AL844140	AL844140 Mus muscu	c 164	19	1.0	137174	8	AP002484	AP002484 Oryza sat
92	20	1.1	233867	2	AC125541	AC125541 Mus muscu	c 165	19	1.0	137571	2	AC079539	AC079539 Mus muscu
93	20	1.1	269576	2	AC073684	AC073684 Mus muscu	c 166	19	1.0	140661	2	AC012425	AC012425 Homo sapi
94	20	1.1	274210	10	AC098732	AC098732 Mus muscu	c 167	19	1.0	141296	2	AL354714	AL354714 Human DNA
95	20	1.1	288050	1	AJ141411	AJ141411 Yersinia	c 168	19	1.0	142961	2	AC109956	AC109956 Rattus no
96	20	1.1	291050	1	AP000982	AP000982 Sulfolobu	c 169	19	1.0	142961	2	AC114236	AC114236 Rattus no
97	20	1.1	349980	6	AX344569	AX344569 Sequence	c 170	19	1.0	143322	2	AC127106	AC127106 Rattus no
98	19	1.0	324	14	BSP277537	277537 Bacterium 2	c 171	19	1.0	143688	2	AC127169	AC127169 Medicago
99	19	1.0	345	14	AF000468	AF000468 HIV-1 str	c 172	19	1.0	144263	2	AP005054	AP005054 Oryza sat
100	19	1.0	516	8	AF441504	AF441504 Pinus taee	c 173	19	1.0	144629	2	AC107254	AC107254 Rattus no
101	19	1.0	1185	9	BC012058	BC012058 Homo sapi	c 174	19	1.0	144784	2	AC009224	AC009224 Homo sapi
102	19	1.0	1685	9	AK024052	AK024052 Homo sapi	c 175	19	1.0	144829	2	AC110571	AC110571 Mus muscu
103	19	1.0	1690	9	AK022763	AK022763 Homo sapi	c 176	19	1.0	145511	2	AC103918	AC103918 Rattus no
104	19	1.0	1749	6	AX374248	AX374248 Sequence	c 177	19	1.0	147974	2	AC076970	AC076970 Homo sapi
105	19	1.0	1940	8	ZMSHO2	AF136221 Zea mays	c 178	19	1.0	148491	9	AC026753	AC026753 Homo sapi
106	19	1.0	2364	9	HS084401	HS084401 Human smoot	c 179	19	1.0	148744	2	AL355576	AL355576 Human DNA
107	19	1.0	2972	6	AR088794	AR088794 Sequence	c 180	19	1.0	148980	2	AC105105	AC105105 Homo sapi
108	19	1.0	2972	6	ARI36665	ARI36665 Sequence	c 181	19	1.0	148980	2	AP002094	AP002094 Oryza sat
109	19	1.0	2972	6	BD005729	BD005729 Vertebrat	c 182	19	1.0	148985	9	AP001172	AP001172 Homo sapi
110	19	1.0	3531	9	AF120103	AF120103 Homo sapi	c 183	19	1.0	149114	8	AC024072	AC024072 Homo sapi
111	19	1.0	3781	9	BC009989	BC009989 Homo sapi	c 184	19	1.0	150165	2	AC094207	AC094207 Rattus no
112	19	1.0	4493	3	MED308548	AJ308548 MYtilus e	c 185	19	1.0	153698	2	AC087554	AC087554 Papio cyn
113	19	1.0	5880	3	CELE7653	X91045 C. elegans 1	c 186	19	1.0	153824	2	AC099055	AC099055 Homo sapi
114	19	1.0	8895	9	HSSMO3	AF114821 Homo sapi	c 187	19	1.0	155962	9	AC092409	AC092409 Papio cyn
115	19	1.0	10029	1	AE013021	AE013021 Thermotana	c 188	19	1.0	156205	2	AC120713	AC120713 Rattus no
116	19	1.0	10752	2	AC097146	AC097146 Rattus no	c 189	19	1.0	156775	2	AC106936	AC106936 Rattus no
117	19	1.0	19082	6	AX345529	AX345529 Sequence	c 190	19	1.0	156949	2	AC018386	AC018386 Homo sapi
118	19	1.0	19413	2	AC014261	AC014261 Drosophi1	c 191	19	1.0	156970	2	AC018386	AC018386 Homo sapi
119	19	1.0	25541	2	AC010174	AC010174 Homo sapi	c 192	19	1.0	157289	2	AC021447	AC021447 Homo sapi
120	19	1.0	29764	3	CEAC8	783087 Caenorhabdi	c 193	19	1.0	157588	9	AL356094	AL356094 Human DNA
121	19	1.0	38414	9	AC098830	AC098830 Homo sapi	c 194	19	1.0	157814	2	AC037460	AC037460 Homo sapi
122	19	1.0	44614	3	CEC29E6	772504 Caenorhabdi	c 195	19	1.0	157911	2	AP005609	AP005609 Oryza sat
123	19	1.0	44830	2	AC014819	AC014819 Drosophi1	c 196	19	1.0	158591	9	AC010188	AC010188 Homo sapi
124	19	1.0	49843	3	AF288092	AF288092 Naeegleria	c 197	19	1.0	158885	9	AC022601	AC022601 Homo sapi
125	19	1.0	60695	2	AC119431	AC119431 Mus muscu	c 198	19	1.0	159336	2	AC020584	AC020584 Homo sapi
126	19	1.0	60695	2	AC119431	AC119431 Mus muscu	c 199	19	1.0	159888	2	AC073164	AC073164 Homo sapi
127	19	1.0	65932	8	AP004938	AP004938 Lotus jap	c 200	19	1.0	160119	2	AC130400	AC130400 Homo sapi
128	19	1.0	67345	3	AC024851	AC024851 Caenorhab	c 201	19	1.0	160596	2	AC123144	AC123144 Lemur cat
129	19	1.0	68145	3	AC024200	AC024200 Caenorhab	c 202	19	1.0	160596	2	AC123144	AC123144 Lemur cat
130	19	1.0	69599	2	AC011865	AC011865 Homo sapi	c 203	19	1.0	160789	2	AL589655	AL589655 Human DNA
131	19	1.0	71892	2	AC129975	AC129975 Mus muscu	c 204	19	1.0	161289	2	AC106906	AC106906 Rattus no
132	19	1.0	73543	2	AC100117	AC100117 Mus muscu	c 205	19	1.0	161938	9	AL360327	AL360327 Human DNA
133	19	1.0	75348	2	AC106315	AC106315 Rattus no	c 206	19	1.0	161943	2	HS84M10	HS84M10 Homo sapi
134	19	1.0	79965	2	AC016223	AC016223 Homo sapi	c 207	19	1.0	162194	2	AC026592	AC026592 Homo sapi
135	19	1.0	81149	9	AC093751	AC093751 Homo sapi	c 208	19	1.0	162224	9	AC012176	AC012176 Homo sapi
136	19	1.0	81301	9	AL159975	AL159975 Human DNA	c 209	19	1.0	162379	2	AC128030	AC128030 Rattus no
137	19	1.0	81565	8	AC006072	AC006072 Arabidops	c 210	19	1.0	162592	2	AC121960	AC121960 Mus muscu
138	19	1.0	83639	9	AL353787	AL353787 Human DNA	c 211	19	1.0	162754	8	OSJN00169	OSJN00169 Oryza sat

c 212	19	1.0 162846	2	AL672275	AL672275 Mus muscu	285	19	1.0 186915	2	AC091574	AC091574 Homo sapi
c 213	19	1.0 163132	9	AC025038	AC025038 Homo sapi	286	19	1.0 187069	2	AC021787	AC021787 Homo sapi
c 214	19	1.0 163279	3	AC104508	AC104508 Drosophil	287	19	1.0 187102	2	AC090217	AC090217 Homo sapi
c 215	19	1.0 163367	2	AC129807	AC129807 Homo sapi	288	19	1.0 187271	9	AC099506	AC099506 Homo sapi
c 216	19	1.0 163421	2	AC012664	AC012664 Homo sapi	c 289	19	1.0 187508	9	AC104078	AC104078 Homo sapi
c 217	19	1.0 163460	2	AC069378	AC069378 Homo sapi	c 290	19	1.0 187649	2	AC068838	AC068838 Homo sapi
c 218	19	1.0 164168	9	AC011930	AC011930 Homo sapi	c 291	19	1.0 187954	2	AC110547	AC110547 Mus muscu
c 219	19	1.0 164448	2	AC102829	AC102829 Mus muscu	c 292	19	1.0 188107	2	AC113229	AC113229 Sus scrof
c 220	19	1.0 164468	2	AC021899	AC021899 Homo sapi	c 293	19	1.0 189605	2	AC122281	AC122281 Mus muscu
c 221	19	1.0 164498	9	AC020898	AC020898 Homo sapi	c 294	19	1.0 189853	2	AC130423	AC130423 Homo sapi
c 222	19	1.0 165124	2	AC110371	AC110371 Rattus no	c 295	19	1.0 191063	9	AC091825	AC091825 Homo sapi
c 223	19	1.0 165126	9	CNS01RHH	Al162511 Huanan chr	c 296	19	1.0 192307	9	AC005971	AC005971 Homo sapi
c 224	19	1.0 165146	2	AP002415	AP002415 Homo sapi	c 297	19	1.0 192729	2	AC110345	AC110345 Rattus no
c 225	19	1.0 165407	2	AC025136	AC025136 Homo sapi	c 298	19	1.0 192951	2	AL671921	AL671921 Mus muscu
c 226	19	1.0 165496	9	AC104461	AC104461 Homo sapi	c 299	19	1.0 193867	9	AC022179	AC022179 Homo sapi
c 227	19	1.0 166039	2	AC044855	AC044855 Homo sapi	c 300	19	1.0 194091	9	AP003534	AP003534 Homo sapi
c 228	19	1.0 166054	9	AC069550	AC069550 Sequence	c 301	19	1.0 194182	2	AC113010	AC113010 Mus muscu
c 229	19	1.0 166061	2	AC096002	AC096002 Rattus no	c 302	19	1.0 194808	2	AC120857	AC120857 Mus muscu
c 230	19	1.0 166100	10	AL645542	AL645542 Mouse DNA	c 303	19	1.0 195002	2	AC123812	AC123812 Mus muscu
c 231	19	1.0 166277	9	AP000852	AP000852 Homo sapi	c 304	19	1.0 195323	2	AP003549	AP003549 Homo sapi
c 232	19	1.0 166345	2	AC068728	AC068728 Homo sapi	c 305	19	1.0 196642	2	AC012587	AC012587 Homo sapi
c 233	19	1.0 166352	2	AC026180	AC026180 Homo sapi	c 306	19	1.0 196788	2	AC021626	AC021626 Homo sapi
c 234	19	1.0 166472	2	AP004753	AP004753 Oryza sat	c 307	19	1.0 197423	2	AC107448	AC107448 Rattus no
c 235	19	1.0 166518	9	AC090514	AC090514 Homo sapi	c 308	19	1.0 197709	9	AC073641	AC073641 Homo sapi
c 236	19	1.0 166604	2	AC068386	AC068386 Homo sapi	c 309	19	1.0 198054	2	AC126995	AC126995 Rattus no
c 237	19	1.0 167758	2	AC131202	AC131202 Rattus no	c 310	19	1.0 198090	2	AC102049	AC102049 Mus muscu
c 238	19	1.0 168893	2	AC032018	AC032018 Homo sapi	c 311	19	1.0 199016	3	AC008204	AC008204 Drosophil
c 239	19	1.0 169372	2	AC079047	AC079047 Homo sapi	c 312	19	1.0 200625	2	AL772157	AL772157 Mus muscu
c 240	19	1.0 169413	2	AC114888	AC114888 Pan trogl	c 313	19	1.0 200774	2	AC118224	AC118224 Mus muscu
c 241	19	1.0 169595	2	AC069316	AC069316 Homo sapi	c 314	19	1.0 200885	2	AC012460	AC012460 Homo sapi
c 242	19	1.0 170367	2	AC025804	AC025804 Homo sapi	c 315	19	1.0 201296	10	AC115299	AC115299 Mus muscu
c 243	19	1.0 170528	9	AC083866	AC083866 Homo sapi	c 316	19	1.0 201723	2	AC112093	AC112093 Rattus no
c 244	19	1.0 170944	9	AC093484	AC093484 Homo sapi	c 317	19	1.0 201913	2	AC094886	AC094886 Rattus no
c 245	19	1.0 170945	9	AC087456	AC087456 Homo sapi	c 318	19	1.0 202134	2	AL772249	AL772249 Mus muscu
c 246	19	1.0 170965	2	AC108288	AC108288 Rattus no	c 319	19	1.0 204442	3	AE003747	AE003747 Drosophil
c 247	19	1.0 171245	9	AC044791	AC044791 Homo sapi	c 320	19	1.0 207982	2	AC121606	AC121606 Mus muscu
c 248	19	1.0 171343	9	AC107030	AC107030 Homo sapi	c 321	19	1.0 212643	2	AC073549	AC073549 Homo sapi
c 249	19	1.0 171482	9	AC034247	AC034247 Homo sapi	c 322	19	1.0 212693	2	AL773516	AL773516 Mus muscu
c 250	19	1.0 171782	2	AC095121	AC095121 Rattus no	c 323	19	1.0 213119	10	AC091694	AC091694 Mus muscu
c 251	19	1.0 172152	2	AC115188	AC115188 Rattus no	c 324	19	1.0 214536	2	AC022810	AC022810 Homo sapi
c 252	19	1.0 172187	9	AC104985	AC104985 Homo sapi	c 325	19	1.0 215253	2	AL845511	AL845511 Mus muscu
c 253	19	1.0 172282	3	AC008366	AC008366 Drosophil	c 326	19	1.0 215295	2	AC112684	AC112684 Mus muscu
c 254	19	1.0 172884	2	AC091164	AC091164 Homo sapi	c 327	19	1.0 215648	2	AC125727	AC125727 Rattus no
c 255	19	1.0 172911	9	AL607026	AL607026 Human DNA	c 328	19	1.0 217615	9	AC011005	AC011005 Homo sapi
c 256	19	1.0 172939	2	AC110959	AC110959 Rattus no	c 329	19	1.0 219744	2	AL773514	AL773514 Mus muscu
c 257	19	1.0 172991	9	AL353608	AL353608 Human DNA	c 330	19	1.0 221377	2	AC097129	AC097129 Rattus no
c 258	19	1.0 174579	2	AC127777	AC127777 Rattus no	c 331	19	1.0 227215	2	AL663092	AL663092 Mus muscu
c 259	19	1.0 175625	2	AC020776	AC020776 Homo sapi	c 332	19	1.0 228502	2	AC111540	AC111540 Rattus no
c 260	19	1.0 175744	2	AC109189	AC109189 Mus muscu	c 333	19	1.0 229164	2	AC107008	AC107008 Rattus no
c 261	19	1.0 175833	9	AC046135	AC046135 Homo sapi	c 334	19	1.0 230850	2	AC112799	AC112799 Rattus no
c 262	19	1.0 176063	2	AC024035	AC024035 Homo sapi	c 335	19	1.0 234346	2	AL645806	AL645806 Mus muscu
c 263	19	1.0 176391	9	AL365400	AL365400 Human DNA	c 336	19	1.0 237727	2	AC099291	AC099291 Rattus no
c 264	19	1.0 176544	2	AL139877	AL139877 Homo sapi	c 337	19	1.0 238267	2	AC096322	AC096322 Rattus no
c 265	19	1.0 176645	9	AC068712	AC068712 Homo sapi	c 338	19	1.0 241543	2	AC091278	AC091278 Mus muscu
c 266	19	1.0 176861	9	AC097065	AC097065 Homo sapi	c 339	19	1.0 246310	2	AC122906	AC122906 Mus muscu
c 267	19	1.0 176937	2	AC126837	AC126837 Rattus no	c 340	19	1.0 248201	2	AC115293	AC115293 Mus muscu
c 268	19	1.0 177602	2	AC025051	AC025051 Homo sapi	c 341	19	1.0 252287	2	AC096323	AC096323 Rattus no
c 269	19	1.0 178062	9	CNS01DRK	Al117356 Human chr	c 342	19	1.0 256591	2	AC115517	AC115517 Rattus no
c 270	19	1.0 178068	2	AP002411	AP002411 Homo sapi	c 343	19	1.0 257445	2	AL845310	AL845310 Mus muscu
c 271	19	1.0 178984	2	AC084023	AC084023 Oryza sat	c 344	19	1.0 280464	2	AC021061	AC021061 Mus muscu
c 272	19	1.0 179085	9	AC011611	AC011611 Homo sapi	c 345	19	1.0 293930	2	AC122220	AC122220 Mus muscu
c 273	19	1.0 179539	2	AC096926	AC096926 Rattus no	c 346	19	1.0 295683	2	AL732498	AL732498 Homo sapi
c 274	19	1.0 179929	2	AC019021	AC019021 Homo sapi	c 347	19	1.0 302040	3	AE003564	AE003564 Drosophil
c 275	19	1.0 179993	9	AC1018360	AC1018360 Homo sapi	c 348	19	1.0 326729	2	AC127952	AC127952 Rattus no
c 276	19	1.0 180078	2	AC120513	AC120513 Papio cyn	c 349	19	1.0 340000	3	AP001666	AP001666 Homo sapi
c 277	19	1.0 181274	2	AC130255	AC130255 Homo sapi	c 350	18	1.0 300	3	AB030084	AB030084 Hydra mag
c 278	19	1.0 181356	2	AC110210	AC110210 Mus muscu	c 351	18	1.0 309	4	SSMSPS12R	SSMSPS12R S. scrofa mi
c 279	19	1.0 182475	2	AC068297	AC068297 Homo sapi	c 352	18	1.0 459	6	AX395802	AX395802 Sequence
c 280	19	1.0 182705	3	AC091208	AC091208 Drosophil	c 353	18	1.0 459	11	AU026501	AU026501 Rattus no
c 281	19	1.0 183522	2	AC121996	AC121996 Mus muscu	c 354	18	1.0 470	3	AF177908	AF177908 Hydra vul
c 282	19	1.0 183559	2	AC129788	AC129788 Rattus no	c 355	18	1.0 487	8	FSU08820	FSU08820 Pisum sativ
c 283	19	1.0 184057	2	AC125301	AC125301 Rattus no	c 356	18	1.0 498	3	AF194965	AF194965 Meloidogy
c 284	19	1.0 185125	2	AC101663	AC101663 Mus muscu	c 357	18	1.0 499	11	G58880	G58880 SHGC-105539

358	18	1.0	636	8	AF153212	Secale ce	431	18	1.0	10882	1	AE007322	AE007322 Streptococ
C 359	18	1.0	665	11	AF1593770	Melaeagris	C 432	18	1.0	10885	1	AE007795	AE007795 Clostridi
360	18	1.0	700	6	AX183253	Sequence	433	18	1.0	11008	3	AE078161	AE078161 Manduca s
361	18	1.0	700	6	AX183254	Sequence	C 434	18	1.0	11540	1	AE007573	AE007573 Clostridi
C 362	18	1.0	922	6	AX281521	Sequence	C 435	18	1.0	12491	1	AE000033	AE000033 Mycoplasma
C 363	18	1.0	922	6	AX347174	Sequence	C 436	18	1.0	13422	1	AE008636	AE008636 Rickettsi
C 364	18	1.0	977	8	AF160305	Arabidops	437	18	1.0	13670	1	U32834	U32834 Haemophilus
C 365	18	1.0	997	8	LEPSBMXR	L.esculentu	438	18	1.0	13944	1	AE000765	AE000765 Aquifex a
C 366	18	1.0	1239	3	AY118665	Arabidops	C 439	18	1.0	14643	1	AE014097	AE014097 Buchnera
367	18	1.0	1241	8	AF479058	Sanderson	440	18	1.0	15704	3	AF067212	AF067212 Caenorhab
368	18	1.0	1315	6	AR052126	Sequence	441	18	1.0	16550	5	PRMTCG	L77119 Methanococc
369	18	1.0	1337	6	AR104312	Sequence	C 442	18	1.0	16646	5	PRMTCG	L42813 Proteobact
370	18	1.0	1337	6	AK026236	Arabidops	443	18	1.0	16764	1	BTB31171	LJ11171 Bacteroid
371	18	1.0	1350	8	AF370322	Arabidops	C 444	18	1.0	16865	2	CER31BE9	Z99785 Caenorhabdi
372	18	1.0	1443	8	TAU07605	Thundergia	445	18	1.0	17952	2	AC014799	AC014799 Drosophila
373	18	1.0	1568	9	BC009822	Sequence	C 446	18	1.0	18025	9	AL663120	AL663120 Human DNA
374	18	1.0	1713	9	AK021486	Homo sapi	C 447	18	1.0	18882	3	D90735	D90735 Escherichia
C 375	18	1.0	1752	9	AK094732	Homo sapi	448	18	1.0	19282	3	J93655	J93655 Caenorhabdi
C 376	18	1.0	1842	8	ATHCDPKA	Arabidops	C 449	18	1.0	19650	1	D90736	D90736 Escherichia
C 377	18	1.0	1895	8	AY070054	Arabidops	C 450	18	1.0	20000	8	SPAPBBB6	AL691490 S.pombe c
C 378	18	1.0	1895	8	AF056493	Pisum sat	451	18	1.0	20705	14	KSU40377	U40377 Kapost's sa
379	18	1.0	1942	14	RFU00761	Rabdit fibr	C 452	18	1.0	20710	6	AR037246	AR037246 Sequence
380	18	1.0	2259	6	AR037256	Sequence	C 453	18	1.0	20710	6	AR119060	AR119060 Sequence
381	18	1.0	2259	6	AR119070	Sequence	454	18	1.0	21131	3	U42838	U42838 Caenorhabdi
382	18	1.0	2327	3	AF305213	Aedes aeg	455	18	1.0	23873	2	AC018333	AC018333 Drosophila
C 383	18	1.0	2381	7	LLKSLV	Insertion s	456	18	1.0	24099	2	AC129023	AC129023 Mus muscu
384	18	1.0	2487	6	AX098604	Sequence	C 457	18	1.0	25906	9	AF178081S1	AF178081 Homo sapi
385	18	1.0	2487	6	AX098606	Sequence	C 458	18	1.0	26241	8	CRU400821	AF178081 Homo sapi
386	18	1.0	2564	6	PATLR3	Sequence	C 459	18	1.0	26802	3	CEFA3D2	CEFA3D2 Homo sapi
387	18	1.0	2583	6	AX034569	Sequence	460	18	1.0	27777	3	AC108926	AC108926 Homo sapi
388	18	1.0	2638	8	PSM3OLD	Sequence	C 461	18	1.0	28026	3	AF125462	AF125462 Caenorhab
C 389	18	1.0	2791	8	PSMWDAG	Sequence	462	18	1.0	28473	6	BD003763	BD003763 Polynucle
390	18	1.0	2902	1	AF270130	Staphyloc	463	18	1.0	30093	3	AF016665	AF016665 Caenorhab
391	18	1.0	2902	6	AX145448	Sequence	464	18	1.0	31021	3	U41277	U41277 Caenorhabdi
392	18	1.0	3037	5	AF393655	Xenopus 1	465	18	1.0	34327	2	AC010426	AC010426 Homo sapi
393	18	1.0	3105	3	AF320764	Spodopter	466	18	1.0	34734	2	AC012844	AC012844 Drosophila
C 394	18	1.0	3226	9	BC015049	Homo sapi	C 467	18	1.0	35100	6	AR065850	AR065850 Sequence
C 395	18	1.0	3236	9	BC015049	Homo sapi	C 468	18	1.0	35100	6	AR127848	AR127848 Sequence
396	18	1.0	3307	1	BCP4821MO	Escherichia	C 469	18	1.0	35100	6	AR194750	AR194750 Sequence
C 397	18	1.0	3329	9	AK095471	Homo sapi	470	18	1.0	35100	6	AR194750	AR194750 Sequence
C 398	18	1.0	3350	6	AF269978	Staphyloc	471	18	1.0	35414	3	AC005328	AC005328 Homo sapi
C 399	18	1.0	3350	6	AX145296	Sequence	472	18	1.0	35422	3	CER09C3	CER09C3 Homo sapi
C 400	18	1.0	3383	9	HSMB03782	Homo sapi	473	18	1.0	36609	3	AF391294	AF391294 Branchios
401	18	1.0	3715	6	AR153661	Sequence	474	18	1.0	38126	2	AC074068	AC074068 Homo sapi
402	18	1.0	3715	6	HSAMPI28	Sequence	C 475	18	1.0	38192	2	AC020177	AC020177 Drosophila
403	18	1.0	3847	10	AB000776	Rattus no	476	18	1.0	40138	6	AR118084	AR118084 Sequence
C 404	18	1.0	3921	8	AF160481	Juncus ef	C 477	18	1.0	40214	3	U32305	U32305 Caenorhabdi
C 405	18	1.0	4239	14	TIVL96	Mouse odc g	C 478	18	1.0	40533	2	AC098758	AC098758 Rattus no
C 406	18	1.0	4260	10	MMODC2	Sequence	479	18	1.0	41633	9	AC096585	AC096585 Homo sapi
C 407	18	1.0	4406	8	CRECPRNLD	Chlamydomon	480	18	1.0	42141	3	CER54G9A	CER54G9A Caenorhab
408	18	1.0	4604	3	DME459772	Drosophila	C 481	18	1.0	42216	3	CER92899	CER92899 Homo sapi
409	18	1.0	4714	9	H8U79734	Human hunti	C 482	18	1.0	42605	3	CRC3BD9	CRC3BD9 Homo sapi
410	18	1.0	4758	5	AF393654	Xenopus 1	483	18	1.0	43746	9	AC093641	AC093641 Homo sapi
411	18	1.0	4796	6	AR153626	Sequence	C 484	18	1.0	45675	3	AF039047	AF039047 Homo sapi
412	18	1.0	4833	8	UPA250427	Uromyces	485	18	1.0	47827	8	AP000731	AP000731 Arabidops
413	18	1.0	4850	8	F24781508	Homo sapi	486	18	1.0	48063	8	AF288090	AF288090 Rhodospira
C 414	18	1.0	5180	9	AF081457	Pisum sat	C 487	18	1.0	48253	9	AL731797	AL731797 Human DNA
C 415	18	1.0	5215	8	AK090416	Homo sapi	C 488	18	1.0	49336	1	AC025950	AC025950 Staphyloc
416	18	1.0	5470	9	HSMB04517	Homo sapi	489	18	1.0	49357	2	AC107961	AC107961 Homo sapi
C 417	18	1.0	5697	1	SYOPSAB	Cyanobacter	C 490	18	1.0	49743	2	AC100574	AC100574 Mus muscu
C 418	18	1.0	5710	6	AX346524	Sequence	491	18	1.0	50152	1	AC025948	AC025948 Staphyloc
419	18	1.0	6564	6	AX345025	Sequence	C 492	18	1.0	50809	2	AL356959	AL356959 Human DNA
420	18	1.0	6756	1	AF012886	Sequence	493	18	1.0	52857	2	AC015159	AC015159 Drosophila
421	18	1.0	6915	1	SMA6085	Sequence	494	18	1.0	53361	5	AL592077	AL592077 Zebrafish
C 422	18	1.0	7100	10	MUSODCC	Mouse onit	495	18	1.0	53952	2	AC101584	AC101584 Mus muscu
423	18	1.0	7635	4	CEFS7E7	Caenorhab	496	18	1.0	54166	2	AC126119	AC126119 Homo sapi
424	18	1.0	8144	6	SSY16039	Sus scrofa	497	18	1.0	54166	2	AC106326	AC106326 Rattus no
425	18	1.0	8144	6	A98022	Sequence 1	498	18	1.0	54859	2	AC084004	AC084004 Homo sapi
426	18	1.0	8255	10	NT1NMRKAN	Plasmid NTP	499	18	1.0	54980	2	AC083777	AC083777 Homo sapi
C 427	18	1.0	8582	10	RATAPC	Rattus norv	C 500	18	1.0	55669	2	AC012694	AC012694 Drosophila
C 428	18	1.0	10029	1	AE005292	Escherich	C 501	18	1.0	55754	2	AC012694	AC012694 Drosophila
429	18	1.0	10098	1	AE000396	Escherich	502	18	1.0	56321	10	AC084214	AC084214 Mus muscu
C 430	18	1.0	10784	1	AE000200	Escherich	503	18	1.0	56538	2	AC092561	AC092561 Clona sav

504	18	1.0	57205	2	AC103947	AC103947 Mus muscu
c 505	18	1.0	58023	2	AC105338	AC105338 Homo sapi
c 506	18	1.0	58210	2	AC131254	AC131254 Homo sapi
c 507	18	1.0	58603	2	AC101496	AC101496 Mus muscu
c 508	18	1.0	58827	2	AC103756	AC103756 Homo sapi
c 509	18	1.0	59538	2	AC012951	AC012951 Drosophil
c 510	18	1.0	61912	2	AC124232	AC124232 Homo sapi
c 511	18	1.0	62649	2	AC022552	AC022552 Homo sapi
c 512	18	1.0	62918	9	AL160289	AL160289 Human DNA
c 513	18	1.0	63267	2	AC126361	AC126361 Homo sapi
c 514	18	1.0	64222	2	AC111583	AC111583 Rattus no
c 515	18	1.0	64392	10	AL1731727	AL1731727 Mouse DNA
c 516	18	1.0	65426	2	AC100815	AC100815 Homo sapi
c 517	18	1.0	65467	2	AC017402	AC017402 Drosophil
c 518	18	1.0	65535	2	AC123918	AC123918 Mus muscu
c 519	18	1.0	66225	2	AC124641	AC124641 Mus muscu
c 520	18	1.0	66237	8	AB016892	AB016892 Arabidops
c 521	18	1.0	66492	2	AC068268	AC068268 Homo sapi
c 522	18	1.0	66668	2	AC102913	AC102913 Mus muscu
c 523	18	1.0	66984	9	AL606753	AL606753 Human DNA
c 524	18	1.0	67183	2	AC099955	AC099955 Mus muscu
c 525	18	1.0	67371	2	AC069043	AC069043 Homo sapi
c 526	18	1.0	67457	2	AC124669	AC124669 Mus muscu
c 527	18	1.0	67745	2	AC116856	AC116856 Mus muscu
c 528	18	1.0	68318	2	AC041015	AC041015 Homo sapi
c 529	18	1.0	68512	2	AC120016	AC120016 Mus muscu
c 530	18	1.0	68662	9	AL133293	AL133293 Human DNA
c 531	18	1.0	68840	9	AC119279	AC119279 Mus muscu
c 532	18	1.0	69303	9	AL357496	AL357496 Human DNA
c 533	18	1.0	69429	2	AC100015	AC100015 Mus muscu
c 534	18	1.0	69947	2	AC102200	AC102200 Mus muscu
c 535	18	1.0	70644	9	AL357563	AL357563 Human DNA
c 536	18	1.0	71848	2	AC105203	AC105203 Homo sapi
c 537	18	1.0	72142	9	AF186189	AF186189 Homo sapi
c 538	18	1.0	72255	2	AC101005	AC101005 Mus muscu
c 539	18	1.0	72383	2	AC101173	AC101173 Mus muscu
c 540	18	1.0	72412	2	AC101594	AC101594 Mus muscu
c 541	18	1.0	73427	9	AC117496	AC117496 Homo sapi
c 542	18	1.0	73520	2	AC129522	AC129522 Homo sapi
c 543	18	1.0	74668	8	AF004481	AF004481 Lotus jap
c 544	18	1.0	75818	9	AC108150	AC108150 Homo sapi
c 545	18	1.0	75867	2	AC124088	AC124088 Mus muscu
c 546	18	1.0	76339	9	AL157694	AL157694 Human DNA
c 547	18	1.0	76375	2	AC017591	AC017591 Drosophil
c 548	18	1.0	77663	2	AC137131_3	Continuation (4 of
c 549	18	1.0	78303	2	AC018243	AC018243 Drosophil
c 550	18	1.0	79248	3	AC004640	AC004640 Drosophil
c 551	18	1.0	79271	3	CSY105EBB	AL132877 Caenorhab
c 552	18	1.0	79431	2	AC098256	AC098256 Rattus no
c 553	18	1.0	80677	8	ATM389	AL022223 Arabidops
c 554	18	1.0	81369	6	AX034587	AX034587 Sequence
c 555	18	1.0	82665	2	AC023204	AC023204 Homo sapi
c 556	18	1.0	82685	9	AL136523	AL136523 Human DNA
c 557	18	1.0	86859	9	AL136523	AL136523 Drosophil
c 558	18	1.0	88839	3	AC084447	AC084447 Caenorhab
c 559	18	1.0	89259	9	AL137076	AL137076 Human DNA
c 560	18	1.0	89469	8	ATF19B3	AL049862 Arabidops
c 561	18	1.0	89665	9	AC063976	AC063976 Homo sapi
c 562	18	1.0	90477	2	AC018551	AC018551 Homo sapi
c 563	18	1.0	91274	8	ATF18N11	AL132953 Arabidops
c 564	18	1.0	91350	9	HSU586J11	AL109828 Human DNA
c 565	18	1.0	91417	2	AC111633	AC111633 Rattus no
c 566	18	1.0	91573	9	AC012372	AC012372 Homo sapi
c 567	18	1.0	91811	10	AL670670	AL670670 Mouse DNA
c 568	18	1.0	91865	9	HS7193H23	AL121367 Human DNA
c 569	18	1.0	92276	2	AC023646	AC023646 Homo sapi
c 570	18	1.0	92496	9	AL589848	AL589848 Human DNA
c 571	18	1.0	92657	8	ATT15N24	AL078465 Arabidops
c 572	18	1.0	92896	9	AL603908	AL603908 Human DNA
c 573	18	1.0	93188	9	AC093674	AC093674 Homo sapi
c 574	18	1.0	93209	2	AC122166	AC122166 Medicago
c 575	18	1.0	93711	8	AC091880	AC091880 Homo sapi
c 576	18	1.0	93724	8	AC007592	AC007592 Genomic s
577	18	1.0	93778	9	AC016999	AC016999 Homo sapi
c 578	18	1.0	93835	2	AC021904	AC021904 Homo sapi
c 579	18	1.0	93900	10	AL645793	AL645793 Mouse DNA
c 580	18	1.0	93950	2	AC130006	AC130006 Rattus no
c 581	18	1.0	95008	2	AC087654	AC087654 Homo sapi
c 582	18	1.0	96100	9	AC092470	AC092470 Homo sapi
c 583	18	1.0	96201	9	AC079601	AC079601 Homo sapi
c 584	18	1.0	96792	9	HSU95741	U95741 Human chrom
c 585	18	1.0	97390	2	AC115490	AC115490 Rattus no
c 586	18	1.0	97614	2	AC096521	AC096521 Rattus no
c 587	18	1.0	98094	2	AC097224	AC097224 Rattus no
c 588	18	1.0	98885	2	AC119624	AC119624 Rattus no
c 589	18	1.0	99077	2	OSJN00262	AL731616 Oryza sat
c 590	18	1.0	99769	2	AC095343	AC095343 Rattus no
c 591	18	1.0	100111	9	AL353679	AL353679 Human DNA
c 592	18	1.0	100678	9	AC109640	AC109640 Homo sapi
c 593	18	1.0	101001	9	AC067735	AC067735 Homo sapi
c 594	18	1.0	101934	2	AC097000	AC097000 Rattus no
c 595	18	1.0	102119	9	AL139005	AL139005 Human DNA
c 596	18	1.0	103004	9	HS171H20	AL031315 Human DNA
c 597	18	1.0	103038	2	AC106701	AC106701 Rattus no
c 598	18	1.0	103177	9	AP000841	AP000841 Homo sapi
c 599	18	1.0	103395	2	AC126792	AC126792 Medicago
c 600	18	1.0	104278	3	AC005269	AC005269 Drosophil
c 601	18	1.0	104363	2	AP002510	AP002510 Homo sapi
c 602	18	1.0	104417	9	AL672220	AL672220 Human DNA
c 603	18	1.0	104802	2	AP003026	AP003026 Homo sapi
c 604	18	1.0	105387	2	AC130236	AC130236 Rattus no
c 605	18	1.0	105565	2	AC104103	AC104103 Mus muscu
c 606	18	1.0	105891	9	AC004850	AC004850 Homo sapi
c 607	18	1.0	106729	9	AC008467	AC008467 Homo sapi
c 608	18	1.0	107132	2	AC096280	AC096280 Rattus no
c 609	18	1.0	107558	2	AC093252	AC093252 Homo sapi
c 610	18	1.0	107689	9	AC010299	AC010299 Homo sapi
c 611	18	1.0	108358	2	AC109904	AC109904 Rattus no
c 612	18	1.0	108479	9	AC068688	AC068688 Homo sapi
c 613	18	1.0	108767	8	AC011809	AC011809 Arabidops
c 614	18	1.0	108902	2	AC011430	AC011430 Homo sapi
c 615	18	1.0	108980	2	AC105210	AC105210 Homo sapi
c 616	18	1.0	109273	2	AC130165	AC130165 Rattus no
c 617	18	1.0	109791	2	AC111400	AC111400 Rattus no
c 618	18	1.0	109933	2	AC093285	AC093285 Homo sapi
c 619	18	1.0	109976	2	AC098548	AC098548 Rattus no
c 620	18	1.0	110000	1	EC00M67_0	U18997 Escherichia
c 621	18	1.0	110000	1	REU0928_1	Continuation (2 of
c 622	18	1.0	110000	2	AC095925_1	Continuation (2 of
c 623	18	1.0	110000	2	AC119565_2	Continuation (3 of
c 624	18	1.0	110000	2	AC120297_2	Continuation (3 of
c 625	18	1.0	110000	2	CEY105E8_1	Continuation (2 of
c 626	18	1.0	110000	2	CEY105E8_2	Continuation (3 of
c 627	18	1.0	110000	2	CEY105E8_5	Continuation (6 of
c 628	18	1.0	110000	2	AC026388_0	AC026388 Mus muscu
c 629	18	1.0	110000	2	AC06388_1	Continuation (2 of
c 630	18	1.0	110004	2	AC106477	AC106477 Rattus no
c 631	18	1.0	110989	2	AC122886	AC122886 Mus muscu
c 632	18	1.0	111357	10	AC110422	AC110422 Mus muscu
c 633	18	1.0	112241	2	AC123561	AC123561 Rattus no
c 634	18	1.0	112659	9	AL137178	AL137178 Human DNA
c 635	18	1.0	113000	9	AP003169	AP003169 Homo sapi
c 636	18	1.0	113218	2	CNS08C8C	AL731872 Oryza sat
c 637	18	1.0	113565	2	AC095567	AC095567 Rattus no
c 638	18	1.0	114041	2	AP001868	AP001868 Homo sapi
c 639	18	1.0	114119	2	AC112288	AC112288 Rattus no
c 640	18	1.0	114242	2	AF236876	AF236876 Homo sapi
c 641	18	1.0	114270	2	AC108194	AC108194 Felis cat
c 642	18	1.0	114713	9	HSJ076301	AL109843 Human DNA
c 643	18	1.0	115641	8	F8K4	AC004392 Arabidops
c 644	18	1.0	115865	9	AL358944	AL358944 Human DNA
c 645	18	1.0	116184	2	AC121241	AC121241 Medicago
c 646	18	1.0	116949	2	AC102768	AC102768 Mus muscu
c 647	18	1.0	117854	9	HSJ885A10	AL117383 Human DNA
c 648	18	1.0	118036	2	AC008198	AC008198 Drosophil
c 649	18	1.0	118195	9	AC007899	AC007899 Homo sapi

C 650	18	1.0	118302	2	AC122647	Rattus no	723	18	1.0	137889	9	AC073269	AC073269 Homo sapi
C 651	18	1.0	118755	9	AC090152	Homo sapi	C 724	18	1.0	138231	9	AC093558	AC093558 Homo sapi
C 652	18	1.0	119012	9	AC106732	Homo sapi	C 725	18	1.0	138740	2	AC004344	AP004344 Oryza sat
C 653	18	1.0	119048	2	AP004167	Oryza sat	C 726	18	1.0	140588	2	AC103082	AC103082 Rattus no
C 654	18	1.0	119334	2	AC105231	Homo sapi	C 727	18	1.0	140714	2	RN374E16	AL603726 Rattus no
C 655	18	1.0	119235	2	AC018187	Homo sapi	C 728	18	1.0	140736	2	AP005618	AP005618 Oryza sat
C 656	18	1.0	119689	2	AC126304	Rattus no	C 729	18	1.0	140999	9	AL3556415	AL3556415 Human DNA
C 657	18	1.0	119707	2	AL773521	Sus scrofa	C 730	18	1.0	141456	9	AL355300	AL355300 Human DNA
C 658	18	1.0	119707	2	AL773521	Sus scrofa	C 731	18	1.0	142102	9	AC107952	AC107952 Homo sapi
C 659	18	1.0	120030	9	AC002075	Homo sapi	C 732	18	1.0	142462	9	AC079238	AC079238 Homo sapi
C 660	18	1.0	120082	9	AC016731	Homo sapi	C 733	18	1.0	142527	9	AL589843	AL589843 Human DNA
C 661	18	1.0	120181	2	AC097290	Rattus no	C 734	18	1.0	142715	9	AC008387	AC008387 Homo sapi
C 662	18	1.0	120269	2	AL356737	Human DNA	C 735	18	1.0	142805	9	AL592525	AL592525 Human DNA
C 663	18	1.0	120527	2	AP005487	Oryza sat	C 736	18	1.0	143370	2	AC130412	AC130412 Homo sapi
C 664	18	1.0	120709	2	AC034185	Homo sapi	C 737	18	1.0	143501	2	AC104383	AC104383 Homo sapi
C 665	18	1.0	121233	2	AC008478	Homo sapi	C 738	18	1.0	143678	2	AC114213	AC114213 Rattus no
C 666	18	1.0	121496	3	AC093391	Homo sapi	C 739	18	1.0	143704	2	AC120899	AC120899 Rattus no
C 667	18	1.0	121720	3	AL591491	Human DNA	C 740	18	1.0	143782	2	AC013808	AC013808 Homo sapi
C 668	18	1.0	122697	3	AC006062	Homo sapi	C 741	18	1.0	143914	2	AC022711	AC022711
C 669	18	1.0	123538	2	AC127929	Rattus no	C 742	18	1.0	144097	2	HS11216	HS11216
C 670	18	1.0	123538	2	AC127929	Rattus no	C 743	18	1.0	144162	2	AC008434	AC008434 Homo sapi
C 671	18	1.0	124497	2	HS232G24	Human DNA	C 744	18	1.0	144201	2	HS230119	HS230119
C 672	18	1.0	124815	2	CNS08C84	Oryza sat	C 745	18	1.0	144218	2	AC111496	AC111496
C 673	18	1.0	125395	2	AP005594	Oryza sat	C 746	18	1.0	144223	2	AP004251	AP004251
C 674	18	1.0	125561	9	AP005120	Homo sapi	C 747	18	1.0	144231	9	AC073571	AC073571
C 675	18	1.0	125805	2	AC095833	Rattus no	C 748	18	1.0	144575	14	YD1293568	YD1293568
C 676	18	1.0	125842	2	AL137850	Human DNA	C 749	18	1.0	144905	10	AL590430	AL590430
C 677	18	1.0	125973	2	AC010353	Homo sapi	C 750	18	1.0	145021	2	AC093257	AC093257 Homo sapi
C 678	18	1.0	126036	9	AC027312	Homo sapi	C 751	18	1.0	145247	2	AC120306	AC120306 Oryza sat
C 679	18	1.0	126054	9	AC018764	Homo sapi	C 752	18	1.0	145642	2	AL845420	AL845420 Danio rer
C 680	18	1.0	126622	2	AC094678	Rattus no	C 753	18	1.0	146022	8	AC090713	AC090713
C 681	18	1.0	126684	2	AC121177	Rattus no	C 754	18	1.0	146040	2	AC129734	AC129734 Rattus no
C 682	18	1.0	126895	2	AL773562	Sus scrofa	C 755	18	1.0	146105	10	AL646049	AL646049 Mouse DNA
C 683	18	1.0	126973	2	AC128776	Rattus no	C 756	18	1.0	146190	2	AC173826	AC173826 Mus muscu
C 684	18	1.0	127202	2	OSJN0272	Human DNA	C 757	18	1.0	146742	2	AL579933	AL579933 Human DNA
C 685	18	1.0	127384	9	AL157823	Human DNA	C 758	18	1.0	146797	2	AC069585	AC069585 Homo sapi
C 686	18	1.0	127790	9	AC008885	Homo sapi	C 759	18	1.0	146836	9	AC019317	AC019317 Homo sapi
C 687	18	1.0	127862	9	AL445430	Human DNA	C 760	18	1.0	146880	2	AL845255	AL845255 Mus muscu
C 688	18	1.0	128119	9	AC009271	Homo sapi	C 761	18	1.0	147029	2	AL137068	AL137068 Human DNA
C 689	18	1.0	128635	9	AC073139	Homo sapi	C 762	18	1.0	147168	2	AC080067	AC080067 Homo sapi
C 690	18	1.0	129360	2	AC117079	Dictyoste	C 763	18	1.0	147318	2	AC079759	AC079759 Homo sapi
C 691	18	1.0	129676	9	AL358113	Human DNA	C 764	18	1.0	147381	2	AC119708	AC119708
C 692	18	1.0	129719	9	AL358113	Human DNA	C 765	18	1.0	147520	2	AP005436	AP005436 Homo sapi
C 693	18	1.0	130513	9	AL356451	Human DNA	C 766	18	1.0	147739	2	AC013514	AC013514 Homo sapi
C 694	18	1.0	130639	9	AL356489	Human DNA	C 767	18	1.0	147820	2	AC013849	AC013849 Homo sapi
C 695	18	1.0	131162	2	AC096067	Rattus no	C 768	18	1.0	148293	2	AC087730	AC087730 Pan trogl
C 696	18	1.0	131248	2	AC114230	Rattus no	C 769	18	1.0	148295	2	AL662913	AL662913 Homo sapi
C 697	18	1.0	131310	2	AC114893	Felis cat	C 770	18	1.0	148298	2	AC123810	AC123810 Mus muscu
C 698	18	1.0	131335	9	AC084740	Homo sapi	C 771	18	1.0	148313	2	AC010930	AC010930 Homo sapi
C 699	18	1.0	132396	9	HS0673D20	Human DNA	C 772	18	1.0	148313	9	AC111321	AC111321 Rattus no
C 700	18	1.0	133108	2	AC034196	Homo sapi	C 773	18	1.0	148679	2	AC111845	AC111845 Rattus no
C 701	18	1.0	133531	2	AC108935	Homo sapi	C 774	18	1.0	148737	2	AC110794	AC110794 Homo sapi
C 702	18	1.0	133567	2	AC097973	Rattus no	C 775	18	1.0	148841	2	AC11845	AC11845 Rattus no
C 703	18	1.0	133846	8	AC099400	Oryza sat	C 776	18	1.0	148883	2	AL355807	AL355807 Human DNA
C 704	18	1.0	133925	9	AC004562	Homo sapi	C 777	18	1.0	149143	9	AL355807	AL355807 Human DNA
C 705	18	1.0	134001	9	AC080094	Homo sapi	C 778	18	1.0	149181	2	AC021813	AC021813 Homo sapi
C 706	18	1.0	134071	9	AC004861	Homo sapi	C 779	18	1.0	149384	2	AC006318	AC006318 Homo sapi
C 707	18	1.0	134305	2	AC068743	Homo sapi	C 780	18	1.0	149581	2	AP005484	AP005484 Oryza sat
C 708	18	1.0	134552	2	AC113887	Rattus no	C 781	18	1.0	149695	2	AC016967	AC016967 Homo sapi
C 709	18	1.0	134687	9	AL445259	Human DNA	C 782	18	1.0	149704	9	AC116310	AC116310 Homo sapi
C 710	18	1.0	135255	2	AC013804	Homo sapi	C 783	18	1.0	149790	2	AC022106	AC022106 Homo sapi
C 711	18	1.0	135497	2	AC102346	Mus muscu	C 784	18	1.0	149824	2	AC009919	AC009919 Homo sapi
C 712	18	1.0	135983	9	HSJ595K12	Human DNA	C 785	18	1.0	149945	2	AC124146	AC124146 Rattus no
C 713	18	1.0	136312	9	HSJ595K16	Human DNA	C 786	18	1.0	150133	2	AC104084	AC104084 Drosophila
C 714	18	1.0	136753	2	AC117193	Mus muscu	C 787	18	1.0	150725	8	SO1400848	SO1400848
C 715	18	1.0	137075	2	AC109074	Rattus no	C 788	18	1.0	150902	2	AC098652	AC098652 Homo sapi
C 716	18	1.0	137154	9	AC019252	Homo sapi	C 789	18	1.0	151945	2	AC100754	AC100754 Homo sapi
C 717	18	1.0	137307	14	XSU75698	Homo sapi	C 790	18	1.0	152050	2	AC129044	AC129044 Rattus no
C 718	18	1.0	137625	9	AC104663	Homo sapi	C 791	18	1.0	152174	2	AC079017	AC079017 Homo sapi
C 719	18	1.0	137684	2	AC094327	Rattus no	C 792	18	1.0	152355	2	AC109839	AC109839 Rattus no
C 720	18	1.0	137721	2	AC113928	Rattus no	C 793	18	1.0	152407	9	AC026007	AC026007 Homo sapi
C 721	18	1.0	137809	2	AC106840	Mus muscu	C 794	18	1.0	152742	2	AC023381	AC023381 Homo sapi
C 722	18	1.0	137809	2	AC106840	Mus muscu	C 795	18	1.0	152927	2	AC119767	AC119767 Rattus no



C 796	18	1.0 153258	2	AC026206	AC026206 Homo sapi	869	18	1.0 160525	2	AC027074	AC027074 Homo sapi
C 797	18	1.0 153263	2	AC112863	Rattus no	C 870	18	1.0 160859	2	AC093190	AC093190 Pan trogl
C 798	18	1.0 153297	2	AC094458	Rattus no	C 871	18	1.0 160929	9	AC008677	AC008677 Homo sapi
C 799	18	1.0 153463	2	AC115155	Rattus no	C 872	18	1.0 161088	2	AC103526	AC103526 Rattus no
C 800	18	1.0 153709	2	AC105682	Rattus no	C 873	18	1.0 161151	9	AC011093	AC011093 Homo sapi
C 801	18	1.0 153715	2	AC105373	Felis cat	C 874	18	1.0 161178	8	AC068892	AC068892 Homo sapi
C 802	18	1.0 153777	2	AP005543	Ryza sat	C 875	18	1.0 161250	8	AC079634	AC079634 Genomic S
C 803	18	1.0 153937	9	AL451075	Human DNA	C 876	18	1.0 161363	10	AL645535	AL645535 Mouse DNA
C 804	18	1.0 154162	2	AC069270	Homo sapi	C 877	18	1.0 161685	2	AC013629	AC013629 Homo sapi
C 805	18	1.0 154178	2	AC128057	Rattus no	C 878	18	1.0 161707	2	AC094976	AC094976 Rattus no
C 806	18	1.0 154180	8	AP000399	Ryza sat	C 879	18	1.0 161732	2	AC128654	AC128654 Homo sapi
C 807	18	1.0 154195	9	AL161773	Human DNA	C 880	18	1.0 161754	10	AL662806	AL662806 Mouse DNA
C 808	18	1.0 154373	10	AC121988	Mus muscu	C 881	18	1.0 161884	2	AC103353	AC103353 Mus muscu
C 809	18	1.0 154472	2	AC019304	Homo sapi	C 882	18	1.0 161983	2	AC127547	AC127547 Mus muscu
C 810	18	1.0 154597	9	AP004295	Homo sapi	C 883	18	1.0 162083	9	AC005962	AC005962 Homo sapi
C 811	18	1.0 154753	2	AC020605	Homo sapi	C 884	18	1.0 162273	2	AL113509	AL113509 Homo sapi
C 812	18	1.0 154796	9	AC007372	Homo sapi	C 885	18	1.0 162422	2	AC124751	AC124751 Mus muscu
C 813	18	1.0 155379	9	HS61668	Human DNA	C 886	18	1.0 162433	2	AC024377	AC024377 Homo sapi
C 814	18	1.0 155491	9	AC068339	Homo sapi	C 887	18	1.0 162730	2	AC026054	AC026054 Homo sapi
C 815	18	1.0 155527	2	AC118811	Rattus no	C 888	18	1.0 162743	4	AC091728	AC091728 Bos tauri
C 816	18	1.0 155637	10	AC111012	Mus muscu	C 889	18	1.0 162782	2	AC129310	AC129310 Mus muscu
C 817	18	1.0 155651	2	AC099632	Mus muscu	C 890	18	1.0 162860	2	AC119097	AC119097 Rattus no
C 818	18	1.0 155724	4	AC091316	Sus scrof	C 891	18	1.0 163075	2	AC107412	AC107412 Rattus no
C 819	18	1.0 155838	2	AC015505	Homo sapi	C 892	18	1.0 163246	9	AC010196	AC010196 Homo sapi
C 820	18	1.0 155881	9	AC002465	Human BAC	C 893	18	1.0 163272	2	AC018496	AC018496 Homo sapi
C 821	18	1.0 155982	2	AC015495	Homo sapi	C 894	18	1.0 163330	2	AC024325	AC024325 Homo sapi
C 822	18	1.0 156165	2	AC021453	Homo sapi	C 895	18	1.0 163331	9	AC092802	AC092802 Homo sapi
C 823	18	1.0 156184	2	AC041018	Homo sapi	C 896	18	1.0 163404	2	AC009614	AC009614 Homo sapi
C 824	18	1.0 156232	9	AC021785	Homo sapi	C 897	18	1.0 163569	2	AC079295	AC079295 Homo sapi
C 825	18	1.0 156266	2	AC118770	Rattus no	C 898	18	1.0 163584	2	AC127843	AC127843 Rattus no
C 826	18	1.0 156324	2	AC120079	Rattus no	C 899	18	1.0 163660	2	AC046165	AC046165 Homo sapi
C 827	18	1.0 156361	2	AC026174	Homo sapi	C 900	18	1.0 163795	9	AP000356	AP000356 Homo sapi
C 828	18	1.0 156503	2	AC112411	Rattus no	C 901	18	1.0 164084	2	AC026580	AC026580 Homo sapi
C 829	18	1.0 156589	2	AC011642	Homo sapi	C 902	18	1.0 164567	3	AC091500	AC091500 Drosophil
C 830	18	1.0 156654	8	AC074283	Ryza sat	C 903	18	1.0 164719	3	AC009374	AC009374 Drosophil
C 831	18	1.0 156823	3	AC007572	Drosophil	C 904	18	1.0 164732	2	AC118946	AC118946 Rattus no
C 832	18	1.0 156989	2	AC126063	Rattus no	C 905	18	1.0 164815	2	AC013668	AC013668 Homo sapi
C 833	18	1.0 157126	2	AC121182	Rattus no	C 906	18	1.0 164857	9	AL161448	AL161448 Human DNA
C 834	18	1.0 157332	9	AC095033	Homo sapi	C 907	18	1.0 164891	9	AC116349	AC116349 Homo sapi
C 835	18	1.0 157365	9	AC025754	Homo sapi	C 908	18	1.0 164928	9	AC087752	AC087752 Homo sapi
C 836	18	1.0 157673	9	AC087555	Papio cyn	C 909	18	1.0 164946	9	AP000553	AP000553 Homo sapi
C 837	18	1.0 157835	3	AC009356	Drosophil	C 910	18	1.0 164973	9	AC087897	AC087897 Homo sapi
C 838	18	1.0 158039	10	AL663110	Mouse DNA	C 911	18	1.0 164974	2	AC019082	AC019082 Homo sapi
C 839	18	1.0 158066	9	AC107307	Homo sapi	C 912	18	1.0 165049	9	AL159995	AL159995 Human DNA
C 840	18	1.0 158249	2	AC121929	Mus muscu	C 913	18	1.0 165062	9	AC087834	AC087834 Pan trogl
C 841	18	1.0 158261	2	AC128506	Rattus no	C 914	18	1.0 165077	2	AC026190	AC026190 Homo sapi
C 842	18	1.0 158348	2	AC105634	Rattus no	C 915	18	1.0 165077	2	AC036190	AC036190 Homo sapi
C 843	18	1.0 158375	2	AC080039	Homo sapi	C 916	18	1.0 165129	9	AP003552	AP003552 Homo sapi
C 844	18	1.0 158588	2	AC092273	Homo sapi	C 917	18	1.0 165165	9	AC095031	AC095031 Homo sapi
C 845	18	1.0 158682	2	AC023083	Homo sapi	C 918	18	1.0 165229	2	AC095327	AC095327 Rattus no
C 846	18	1.0 158703	2	AC115970	Mus muscu	C 919	18	1.0 165314	2	AC103081	AC103081 Rattus no
C 847	18	1.0 159070	2	AC131160	Homo sapi	C 920	18	1.0 165720	2	AC025524	AC025524 Homo sapi
C 848	18	1.0 159145	2	AC027172	Homo sapi	C 921	18	1.0 165811	2	AC127974	AC127974 Rattus no
C 849	18	1.0 159232	2	AC128676	Homo sapi	C 922	18	1.0 166035	2	AC115456	AC115456 Rattus no
C 850	18	1.0 159358	2	AC111732	Rattus no	C 923	18	1.0 166096	2	AC120947	AC120947 Rattus no
C 851	18	1.0 159391	2	AC027113	Homo sapi	C 924	18	1.0 166141	9	AL590627	AL590627 Human DNA
C 852	18	1.0 159465	2	AP005001	Ryza sat	C 925	18	1.0 166145	2	AC017064	AC017064 Homo sapi
C 853	18	1.0 159477	2	AC116526	Mus muscu	C 926	18	1.0 166239	2	AC078965	AC078965 Homo sapi
C 854	18	1.0 159482	2	AL512371	Homo sapi	C 927	18	1.0 166266	2	AP001146	AP001146 Homo sapi
C 855	18	1.0 159527	2	AC128826	Rattus no	C 928	18	1.0 166306	9	AC022880	AC022880 Homo sapi
C 856	18	1.0 159563	9	AC023200	Homo sapi	C 929	18	1.0 166317	9	AL157895	AL157895 Human DNA
C 857	18	1.0 159607	2	AC027334	Homo sapi	C 930	18	1.0 166377	2	AC102193	AC102193 Mus muscu
C 858	18	1.0 159610	2	AC012137	Homo sapi	C 931	18	1.0 166447	9	AC018751	AC018751 Homo sapi
C 859	18	1.0 159616	2	AC128311	Rattus no	C 932	18	1.0 166511	9	AC016165	AC016165 Homo sapi
C 860	18	1.0 159716	2	AC025807	Homo sapi	C 933	18	1.0 166520	2	AC121431	AC121431 Rattus no
C 861	18	1.0 159724	2	AC127800	Rattus no	C 934	18	1.0 166555	3	AC010709	AC010709 Drosophil
C 862	18	1.0 159772	9	AC008990	Homo sapi	C 935	18	1.0 166677	3	AC115482	AC115482 Drosophil
C 863	18	1.0 159823	2	AC109961	Rattus no	C 936	18	1.0 166702	9	AP012594	AP012594 Homo sapi
C 864	18	1.0 159857	14	AF170722	Rabbit fi	C 937	18	1.0 166743	9	AP003357	AP003357 Homo sapi
C 865	18	1.0 159882	9	AP003695	Homo sapi	C 938	18	1.0 166819	9	AC008109	AC008109 Homo sapi
C 866	18	1.0 160082	2	AC123125	Rattus no	C 939	18	1.0 166857	2	AC010020	AC010020 Drosophil
C 867	18	1.0 160245	2	AL158031	Homo sapi	C 940	18	1.0 167025	2	AC021380	AC021380 Homo sapi
C 868	18	1.0 160408	2								

c 942	18	1.0	167052	2	AC128655	Homo sapi
c 943	18	1.0	167033	2	AC094951	Rattus no
944	18	1.0	167164	2	AC026506	Homo sapi
945	18	1.0	167231	2	AC115367	Rattus no
946	18	1.0	167302	2	AC099756	Homo sapi
c 947	18	1.0	167398	9	CNS07EET	AL445886 Human chr
948	18	1.0	167525	9	AC006531	Homo sapi
c 949	18	1.0	167568	2	AC111734	Rattus no
950	18	1.0	167749	2	AC073142	Homo sapi
c 951	18	1.0	167753	9	AC108752	Homo sapi
952	18	1.0	167827	2	AC022749	Homo sapi
953	18	1.0	167863	2	AC121949	Mus muscu
c 954	18	1.0	168056	2	AC004458	Homo sapi
955	18	1.0	168066	2	AC026902	Homo sapi
c 956	18	1.0	168097	2	AC102401	Mus muscu
957	18	1.0	168253	2	AC116818	Mus muscu
958	18	1.0	168319	9	AC023282	Homo sapi
959	18	1.0	168332	2	AC053483	Homo sapi
c 960	18	1.0	168357	9	AC104440	Homo sapi
961	18	1.0	168433	2	AC110696	Rattus no
c 962	18	1.0	168478	2	AL611925	Homo sapi
c 963	18	1.0	168483	2	AC094611	Rattus no
964	18	1.0	168505	2	AC080010	Homo sapi
965	18	1.0	168644	9	AC004544	Homo sapi
966	18	1.0	168677	9	CNS0000F	AL049837 Human chr
c 967	18	1.0	168818	2	AC013496	Homo sapi
968	18	1.0	168822	9	AC007511	Homo sapi
969	18	1.0	168946	2	AC125639	Rattus no
c 970	18	1.0	168975	2	AC122511	Mus muscu
c 971	18	1.0	169030	2	AC125625	Homo sapi
972	18	1.0	169106	2	AC120743	Rattus no
c 973	18	1.0	169229	2	AC125913	Rattus no
974	18	1.0	169237	9	AC009516	Homo sapi
c 975	18	1.0	169248	2	AC129338	Homo sapi
c 976	18	1.0	169408	9	AC079258	Homo sapi
977	18	1.0	169537	2	AC107329	Rattus no
c 978	18	1.0	169542	2	AL356482	Homo sapi
979	18	1.0	169572	2	AC023832	Mus muscu
c 980	18	1.0	169607	2	AC023569	Homo sapi
c 981	18	1.0	169660	9	AL161449	Human DNA
c 982	18	1.0	169670	9	AL390962	Human DNA
c 983	18	1.0	169814	2	AC130759	Rattus no
c 984	18	1.0	169877	3	AC104612	Drosophi
c 985	18	1.0	170008	9	AC117391	Homo sapi
986	18	1.0	170145	2	AC113969	Mus muscu
987	18	1.0	170215	9	AC092691	Homo sapi
c 988	18	1.0	170241	2	AC120635	Rattus no
c 989	18	1.0	170251	2	AC009935	Homo sapi
c 990	18	1.0	170291	2	AC106944	Rattus no
991	18	1.0	170316	2	AC122151	Mus muscu
c 992	18	1.0	170332	2	AC106104	Rattus no
c 993	18	1.0	170393	9	AC026473	Homo sapi
994	18	1.0	170414	2	AC026019	Homo sapi
c 995	18	1.0	170472	2	AC125911	Rattus no
c 996	18	1.0	170512	2	AC021806	Homo sapi
997	18	1.0	170578	2	AC110703	Rattus no
998	18	1.0	170608	2	AL359372	Homo sapi
999	18	1.0	170749	2	AC112890	Rattus no
c1000	18	1.0	170772	2	AC014654	Homo sapi

## ALIGNMENTS

RESULT 1  
LOCUS AB001641 14134 bp DNA linear BCT 01-DEC-2000  
DEFINITION Chlamydia pneumoniae section 57 of 103 of the complete genome.  
ACCESSION AB001641 AB001643  
VERSION AB001641.1 GI:4376855

KEYWORDS Chlamydia pneumoniae CWL029.  
SOURCE Chlamydia pneumoniae  
ORGANISM Chlamydia pneumoniae CWL029

```

REFERENCE
  1 (bases 1 to 14134)
    Kalkan, S., Mitchell, W., Marathe, R., Lamell, C., Fan, J., Hyman, R.W.,
    Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.
    Comparative genomes of Chlamydia pneumoniae and C. trachomatis
    Nat. Genet. 21 (4), 385-389 (1999)
    99206606
    MEDLINE
    PUBMED
  2 (bases 1 to 14134)
    Kalkan, S., Mitchell, W., Marathe, R., Lamell, C., Fan, J., Olinger, L.,
    Grimwood, J., Davis, R.W. and Stephens, R.S.
    Direct Submission
    Submitted (01-DEC-1998) Program in Infectious Diseases, University
    of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
    Location/Qualifiers
      source
        organism="Chlamydia pneumoniae CWL029"
        strain="CWL029"
        db_xref="taxon:115713"
        complement(158, .1924)
        gene="recJ"
        note="CpN0563"
        complement(158, .1924)
        gene="recJ"
        /codon_start=1
        /transl_table=11
        /product="ssDNA Exonuclease"
        /protein_id="AAD18703.1"
        /db_xref="GI:4376856"
        /translation="MTNSDNMSAAGLWHPKEDPAFGIMTIKEFLHPPTVAOIFSR
        GROTIOETHEKFLYSLSLYDGLGDSKAEIRLADREHEMHTIGDSNDVMTG
        VALVEPLRIDVHSYIFPLGAIKQHGSTSLIKLKEBGTLLITVDCGTACKGVN
        SDITQIGDIVIITDHHMPGPKIPIHCVATLNPYLRRHTYNNRLETVGVAFKLARGVNL
        ALISNLVPSQSGSLKLDLVTLLGTDVGLDGNVWVRVYKELARGPGLNK
        LCAICGVKSEVSTSDIYLIKAPKNSLGRDDPAKGVALLTODDERVADALIMELN
        INRERORIEAEVFOOVSEIINPEILNOAIVLSTMAHVAIPIISARLAKTYNKP
        VITIAIORIGIGKSARITGSPFLNLKVCSSLISXGHDAAVIAVIMKEDVDEPKK
        KRVHLNSSLKKKGDPLPHLEIDAYVDPAIDYDLASBELPFPCKGNLMKPIFYSKVR
        DVRYPRKVLPGNHLKYLTSQKEMNLGVAFGLGRHADLAKASWHPLEIAYTRLSQTS
        GSGVTHLVRDPRISSEPRFSD"
        complement(2010, .6218)
        /gene="secD/secF"
        /note="CpN0564"
        complement(2010, .6218)
        /gene="secD/secF"
        /codon_start=1
        /transl_table=11
        /product="Protein Export Proteins SecD/SecF (fusion)"
        /protein_id="AAD18704.1"
        /db_xref="GI:4376857"
        /translation="MKQVKNRFAIILCVFALALYYVPTCLYAKPLDKIDGNBAE
        HIKSFTQAOQVRDVIPIVSAIISLHRGHIOHPAIPDIVRFRKGDADPFI
        GNLVGGEVNPVAKTSRPHLVGSRHDHVIQVASSINTSVLSPFSFVSSENEQE
        NASILORVYACFPKOKDCGSPSIVMETPKRLOLVAKNSLGSFVFSRSLAF
        COOFSRSDRLAFLSRLSSNDAAIVDEOKLKSVELTSLQACIRSLDCCYTG
        LKDCBSGLPFSSTIEYCPKKEKTLTTHSDILADRTSISKEQRLDPFRLAVEKQKL
        SKNLTVQVEDVYNGFSFQWMDQVTKGKIILOSERLLOGIAEHLTLTHRPAESCDL
        IPENPFVPCORPSEBAFCYIFSPNTCKKSKSVYILGLGSLVAKYXQOOGAKFL
        LOSPEKDQNLNCFSSHTEALISWTLGEOVLEIRHPLQOFLDVMGEFVYIGKGCAPL
        EYKDIODRLATNOIEIKRQSDVLVHMGHGRKAKSMDOERLSAIPQNLFLKMK
        LMRKPSRGENTLRKIDPVGROLLISFKHQGQOLNDKIDLIVSDELTAKG
        VSEIELRBEGDYLHSVQSSSTSSSEITLGSFKSFHVNNRFSYSASRYVORFLD
        YLMTFSQAQKTSPEIINTFAPALFNEEDVPSVHEAIITKLKSLGLAFSPSGCTPS
        TDLDTTFSMAIGKDAEQKANPLVIFPNRYALDGLASLIDREFPAGGVYLVNFVKD
        TSPKMAEKLSTPESFHTWTSAYCOEGISGTANGQYSANRGRMAVVDIGVWSSPIL
        NVPLKNHVASGSKFTHREVSFKASDLKSGAMSFVPEVISEETISDLDGKCTOGIIS
        ACCGAMAILVMSVYIRPGCVIASCAYVILNLLIMAAIOYDAPLTTISLAIYVAMG
        NADVANVLFERIREEPFLSGLKSVKGVKAKQAIQFDSNLTIVLASALPIFLDTG
        PIKGFALTLITGIFSSMFTALFMTKFPMLMNNKQOHLHMMNFVIGIKDFLPGCK
        KLMAVSGSVFLGCVAGAMNSVYLGDFKGVAFITPNPKKHGSDVAQMGKVVHK
        LOEAGLSGDRPFIOTFGSSSEKIKIYFSDKALSYTADTSLSPKIDHELAALGVGLSE
        TGLDSTETLNTONFWGSKSKSKMKRYOATIGLLALAIILLYVSLRPFMOYAFS
        AVCALIHDLATCAVLPFAHPFKTKIQIDLOAIGALMTVYGLSLNNTLIIIPRIEDR

```

```
gene
/translation="OANLFTPMHVLNDALOKTFSRTVMYTTATTLSVLMLLFIGGSSVFNFAFIMTIGILL
GTLSLYATAPLLLPWYKRNRSK"
complement (6398..7498)
/gene="Cpno565"
CDS
complement (6398..7498)
/gene="Cpno565"
/codon_start=1
/transl_table=11
/product="CT449 hypothetical protein"
/protein_id="AAD18705.1"
/db_xref="GI:4376858"
/translation="MKFPRISIDLIPTQVIMWRGGNVHYVPAONLKKILGGVYL
ACFGLALLGCAFAAGVQOTIFPCIGILMILGLVLLGPAYLQYSGWGRFPRPLRETK
VFEPKINWGLCSLLQSKWKIRPCGYHPGCPQVEICEGQEIWKIFQKSRDNTSI
FLIEMQIALRQIESKSRKTFAPDPSVSSLLSEIQREEQQYIDPKVSWSSSD
PISQDTHPKAEIYVNIISDAQEPQRCYIDAYKAFFTVLQDQDPNIVKKHTIYVLT
PLIGVPALPKEQENIKLASQAFLYSAEQVAKRMREKQDSIRIKFIFTDPTSPS
LYSPHSHSTPHSVTPISLSGFGVEQESYTPA"
8009..8761
/gene="yaes"
/note="Cpno566"
8009..8761
/gene="yaes"
/codon_start=1
/transl_table=11
/product="yaes family hypothetical protein"
/protein_id="AAD18706.1"
/db_xref="GI:4376859"
/translation="MSLATNNAESKFPSSLQRLPNHVAIIMDGNRRWYKKHRECGHTH
TSGHYGAQVLPNTLNAVLDLGKIVLYTFSTENFCRKEEIQEIENIYVTDLDKOL
PYLMENECILRCIGDLSKLPKGIQTKINHVRMTASFRLSLVLAVNYSKDELVRAP
KLUHVDLNLKISDDSESLSISSYLTSTGLTDPDLIRITGGEMRWVSNFLWQIAYTE
LYITDLWDPDTPODLFEAINVYQQRSRGCK"
8762..9688
/gene="cdaA"
/note="Cpno567"
8762..9688
/gene="cdaA"
/codon_start=1
/transl_table=11
/product="Phosphatidate Cytidylyltransferase"
/protein_id="AAD18707.1"
/db_xref="GI:4376860"
/translation="MLNSNFKSKTGAYGBLQFQVVVHSLVLFVLVLLLYSSLPFLTS
FALGFIATCAGVTGYEYSMAKAMHYPILSTFSAIGSFILFALSIFSIWHSLSPGF
FDALPWTLLVWVMSIFVRKSTIGALQSGVTLFSLIYVGPILRFLHVLVSFIHT
QEPVLGIWASFLIATTKGADIFGYFFGKAFGNKKIAPQISPNKTVVGVAGCLGATL
ISPIFFLIQIPTRASVFPMPAILIPLGLALGITGFFGDIIESIFKRDHLKNSNKLKA
VGGMLDPLDSILLSTPIAYLFLITQSKKEFIG"
9685..10335
/gene="cmk"
/note="Cpno568"
9685..10335
/gene="cmk"
/codon_start=1
/transl_table=11
/product="CMP kinase"
/protein_id="AAD18708.1"
/db_xref="GI:4376861"
/translation="MIITIDGSGTGTAKALADHLHFNHCNTGKMYRTLAYRLQ
SPWATPLTLFLEPPSPFTFATGQPLESFNGHLTSELTTQEVANAASELSOLPEV
RAFMDIQRYAQLGNCVFFGRDMGKVFNPADLKIFLTSSPEVRAQRRLKDPGL
SPQLOAELVLRDAADAQRAHDPVLPENGIVIDSSDLTIQVLEKILALLFRNEL"
10332..10970
/gene="plec"
/note="Cpno569"
10332..10970
/gene="plec"
/codon_start=1
/transl_table=11
/product="Glycerol-3-P Acyltransferase"
/protein_id="AAD18709.1"
/db_xref="GI:4376862"
gene
/translation="MIFRICKFTVAFSLPYKLKVVGVKNFIKGPAILAVNHNSL
DPIALHMCVHECIYHLARSLFNIPLMWKQWCFPPQRODEGNSAAFKIARSLFNKRKK
LVYIPEGARSPDQOLQPKVIGMAAKSRVPIIPVYIRGTFFAENRHKIPIPHVMTKI
TCVFGTPMYFDDIIQNPENIKNKETYQIITNQTWNKIAELKAWVESGCKGDP"
10975..12660
/gene="argS"
/note="Cpno570"
10975..12660
/gene="argS"
/codon_start=1
/transl_table=11
/product="Arginyl tRNA Transferase"
/protein_id="AAD18710.1"
/db_xref="GI:4376863"
/translation="NSTLSILSVISQAIKAFPNLDEWAPBITSTKEHFHGYOQN
DAKLARLVKAPRAIEAIVAEIQLPQEFSLIEIAGAGFINFTFSPVFLNQLEHFKD
ALMKLQVQSPKKIIFDPSNIAKAMHVGHLRSTIIGDSLARIFSVYGHVLRNLHI
GDMGTAFGLMITVLOENPCDYSLEDLTSLYKAYVCFNDEBEFKKESQONVALQAK
DPOAIAIWEKICETSEKAFQKIYDILDIIVVEKKGESFYNPFLEIIEDESKGLLTVS
NDACVDFEAFSIPFMVQSGDGYNTATDLAAMRYRIEBEDHADKIIIVTDLQOSLHF
QLLEATAIAAGYLQPGIFSHVGFGLVLDPQKKLKTSGENVKLRRELLDTAIEKAEA
LREHRELPIDAEIQERAPVIGINAIKYSDLSHRTSDYVSEFKMLRFEGNTAMFLLY
AVVIOGIKRRLGISOLSEGPPEIOEPAEELIALTLRPEALESTIKELCPHFLTD
YLYNLTHKNGFPRDSHIQDSPYAKSRLFLCALAEQVLATGMHLGLKTLERL"
complement (12614..13990)
/gene="murA"
/note="Cpno571"
complement (12614..13990)
/gene="murA"
/codon_start=1
/transl_table=11
/product="UDP-N-Acetylglucosamine Transferase"
/protein_id="AAD18711.1"
/db_xref="GI:4376864"
/translation="MOIAQVFCGRLNGEVKUSGAKNAATKLIVASILSDOKCTLRNV
PDIGCVSLVELCKSLGANVSWDKETEVELEIYTPQCTRVPTTFSNVNRPIPLLGA
LIGRCPEGVTVPTGGDAGERTLNFHFEGLKOLGOVQISSDSSGYAKAPRGLKNGVI
HLPVSGVATENLILAAIHAKRTVINKVALEAIBLDLVLFQAGADITDNDRTID
IFQGTGLGSDVHTILPDKIEAASFMAAVVSGRVFVNAKQELLIPFLKMLRSIGCG
FLVSESGIIEFFQERPLVGVVLETDVHPGFLTDWQFPFVALLSQAGSSVIHETVHEN
RLGYLHLOHMGAECCOLFHQCLSTKACRVAINGNFPHSAVTHGATPLWASHLVIPLDFA
GFAYVMAALIAEGGSGSIIENTHLLDRGYNNVWVKLSLGNKIQIDMEQSELTSPKS
LAURASL"
BASE COUNT 4180 a 2981 c 2805 g 4168 t
ORIGIN
Query Match 78.0%; Score 1453; DB 1; Length 14134;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 1; Indels 3; Gaps 3;
QY 1 ATGGACTTCGCGCATATTGTTCAGGAGGGGATCAGCGGCATCTGCTAATGGACAATATTCCTGC 60
DB 3771 ATGGACTTCGCGCATATTGTTCAGGAGGGGATCAGCGGCATCTGCTAATGGACAATATTCCTGC 3712
QY 61 AAACCGTGGATGGCGTATGCTGTAGTGATTCAGCGTTATATGTCAGCAGCCCTATTTT 120
DB 3711 AAACCGTGGATGGCGTATGCTGTAGTGATTCAGCGTTATATGTCAGCAGCCCTATTTT 3652
QY 121 AAACGTCCTCCATTGAAAATCATGCCAGTGTCTCAGGAGAAATTTTACCACCGTGAAGTGAG 180
DB 3651 AAACGTCCTCCATTGAAAATCATGCCAGTGTCTCAGGAGAAATTTTACCACCGTGAAGTGAG 3592
QY 181 CAAACTCGCCTCAGATTAAATACTCGAGCGATGCTCTTTTGTTCGCCAGGTTCTCAGTGA 240
DB 3591 CAAACTCGCCTCAGATTAAATACTCGAGCGATGCTCTTTTGTTCGCCAGGTTCTCAGTGA 3532
QY 241 AGAGACGATCTCTTCTGATCTTGGGAAAAAACAATGTACACAAGGCATTATCTCAGCATG 300
DB 3531 AGAGACGATCTCTTCTGATCTTGGGAAAAAACAATGTACACAAGGCATTATCTCAGCATG 3472
QY 301 CTGTGGCTTCGCAATGCTTATTTTGTGATGAGCGTATATATAGATTTCGAGCGCTCAT 360
DB 3471 CTGTGGCTTCGCAATGCTTATTTTGTGATGAGCGTATATATAGATTTCGAGCGCTCAT 3412
```

```
gene
/translation="OANLFTPMHVLNDALOKTFSRTVMYTTATTLSVLMLLFIGGSSVFNFAFIMTIGILL
GTLSLYATAPLLLPWYKRNRSK"
complement (6398..7498)
/gene="Cpno565"
CDS
complement (6398..7498)
/gene="Cpno565"
/codon_start=1
/transl_table=11
/product="CT449 hypothetical protein"
/protein_id="AAD18705.1"
/db_xref="GI:4376858"
/translation="MKFPRISIDLIPTQVIMWRGGNVHYVPAONLKKILGGVYL
ACFGLALLGCAFAAGVQOTIFPCIGILMILGLVLLGPAYLQYSGWGRFPRPLRETK
VFEPKINWGLCSLLQSKWKIRPCGYHPGCPQVEICEGQEIWKIFQKSRDNTSI
FLIEMQIALRQIESKSRKTFAPDPSVSSLLSEIQREEQQYIDPKVSWSSSD
PISQDTHPKAEIYVNIISDAQEPQRCYIDAYKAFFTVLQDQDPNIVKKHTIYVLT
PLIGVPALPKEQENIKLASQAFLYSAEQVAKRMREKQDSIRIKFIFTDPTSPS
LYSPHSHSTPHSVTPISLSGFGVEQESYTPA"
8009..8761
/gene="yaes"
/note="Cpno566"
8009..8761
/gene="yaes"
/codon_start=1
/transl_table=11
/product="yaes family hypothetical protein"
/protein_id="AAD18706.1"
/db_xref="GI:4376859"
/translation="MSLATNNAESKFPSSLQRLPNHVAIIMDGNRRWYKKHRECGHTH
TSGHYGAQVLPNTLNAVLDLGKIVLYTFSTENFCRKEEIQEIENIYVTDLDKOL
PYLMENECILRCIGDLSKLPKGIQTKINHVRMTASFRLSLVLAVNYSKDELVRAP
KLUHVDLNLKISDDSESLSISSYLTSTGLTDPDLIRITGGEMRWVSNFLWQIAYTE
LYITDLWDPDTPODLFEAINVYQQRSRGCK"
8762..9688
/gene="cdaA"
/note="Cpno567"
8762..9688
/gene="cdaA"
/codon_start=1
/transl_table=11
/product="Phosphatidate Cytidylyltransferase"
/protein_id="AAD18707.1"
/db_xref="GI:4376860"
/translation="MLNSNFKSKTGAYGBLQFQVVVHSLVLFVLVLLLYSSLPFLTS
FALGFIATCAGVTGYEYSMAKAMHYPILSTFSAIGSFILFALSIFSIWHSLSPGF
FDALPWTLLVWVMSIFVRKSTIGALQSGVTLFSLIYVGPILRFLHVLVSFIHT
QEPVLGIWASFLIATTKGADIFGYFFGKAFGNKKIAPQISPNKTVVGVAGCLGATL
ISPIFFLIQIPTRASVFPMPAILIPLGLALGITGFFGDIIESIFKRDHLKNSNKLKA
VGGMLDPLDSILLSTPIAYLFLITQSKKEFIG"
9685..10335
/gene="cmk"
/note="Cpno568"
9685..10335
/gene="cmk"
/codon_start=1
/transl_table=11
/product="CMP kinase"
/protein_id="AAD18708.1"
/db_xref="GI:4376861"
/translation="MIITIDGSGTGTAKALADHLHFNHCNTGKMYRTLAYRLQ
SPWATPLTLFLEPPSPFTFATGQPLESFNGHLTSELTTQEVANAASELSOLPEV
RAFMDIQRYAQLGNCVFFGRDMGKVFNPADLKIFLTSSPEVRAQRRLKDPGL
SPQLOAELVLRDAADAQRAHDPVLPENGIVIDSSDLTIQVLEKILALLFRNEL"
10332..10970
/gene="plec"
/note="Cpno569"
10332..10970
/gene="plec"
/codon_start=1
/transl_table=11
/product="Glycerol-3-P Acyltransferase"
/protein_id="AAD18709.1"
/db_xref="GI:4376862"
```

QY 361 CGCTTGGGAGCTGTTCTTGAATCTTTGCTTATCTGGGACCTTACAGTATTTGGA 420  
 DB 3411 CGCTTGGGAGCTGTTCTTGAATCTTTGCTTATCTGGGACCTTACAGTATTTGGA 3352  
 QY 421 TGGGCCATCAGCTGTCAGGACTGGCTGGGATTTGTTCTTGCTATGAGGAGATGGCCGTAGA 480  
 DB 3351 TGGGCCATCAGCTGTCAGGACTGGCTGGGATTTGTTCTTGCTATGAGGAGATGGCCGTAGA 3292  
 QY 481 TGGCAATGTTCTTGATTCGAAAGAAATCCGAGAGAAATTTTATTTGTCGAAAGCTTAA 540  
 DB 3291 TGGCAATGTTCTTGATTCGAAAGAAATCCGAGAGAAATTTTATTTGTCGAAAGCTTAA 3222  
 QY 541 AAAATCTGAGAAAAAGAAATACCAAGGCTTTTGAGGCCATTTTGATTTCACTTAC 600  
 DB 3231 AAAATCTGAGAAAAAGAAATACCAAGGCTTTTGAGGCCATTTTGATTTCACTTAC 3172  
 QY 601 TACAGTATGGCCCTCAGCACTTCTTTCTTCTAGATACAGGGGCTATTAAGGGTTTGC 660  
 DB 3171 TACAGTATGGCCCTCAGCACTTCTTTCTTCTAGATACAGGGGCTATTAAGGGTTTGC 3112  
 QY 661 TTTGACATTTGATTTAGGAATTTTCTCTCAATGTTTACGGCTTTTCAATGACTAAAT 720  
 DB 3111 TTTGACATTTGATTTAGGAATTTTCTCTCAATGTTTACGGCTTTTCAATGACTAAAT 3052  
 QY 721 TTTCTCATGCTGTGATGATATAGAACCCCAATACACAGTTGCAATGATGAATTAAGTT 780  
 DB 3051 TTTCTCATGCTGTGATGATATAGAACCCCAATACACAGTTGCAATGATGAATTAAGTT 2992  
 QY 781 CGTGGGATTAAGATATTTCTTGAGAGGATGCAAAAACTTTGGGCTGTTTCTGAGAG 840  
 DB 2991 CGTGGGATTAAGATATTTCTTGAGAGGATGCAAAAACTTTGGGCTGTTTCTGAGAG 2932  
 QY 841 TGTCTTTCTTTAGAGTCCCTTGTCTCTGGGTTTGGAGCCTGGAATTCGGTTTGGGAAT 900  
 DB 2931 TGTCTTTCTTTAGAGTCCCTTGTCTCTGGGTTTGGAGCCTGGAATTCGGTTTGGGAAT 2872  
 QY 901 GGATTTTAAAGAGAGGATATCTTTTACCTTTAATCCAAAGAGATGGCATCAGCGATGT 960  
 DB 2871 GGATTTTAAAGAGAGGATATCTTTTACCTTTAATCCAAAGAGATGGCATCAGCGATGT 2812  
 QY 961 TGTCTCAATGCTGTGGCAAACTTGCAATTAACAGGAGCTGTCTTCTTTAGAGA 1020  
 DB 2811 TGTCTCAATGCTGTGGCAAACTTGCAATTAACAGGAGCTGTCTTCTTTAGAGA 2752  
 QY 1021 CTTCGATTCGAACATTTGATCTTCCGAAAGATCAAAATCTATTTAGTGAATAAGC 1080  
 DB 2751 CTTCGATTCGAACATTTGATCTTCCGAAAGATCAAAATCTATTTAGTGAATAAGC 2692  
 QY 1081 TTTAAGCTATACT AAGCAGATACGAGCTCTCTCTAATAATTAACATCATGAGCTGGC 1139  
 DB 2691 TTTAAGCTATACTAAGCAGATACGAGCTCTCTCTAATAATTAACATCATGAGCTGGC 2632  
 QY 1140 GTTA-TTGTGGGATTTGTTGTCAGAAACAGGCTAGATTTCTCTAGGAAATCTTAAAGC 1198  
 DB 2631 GTTAAGCTGTGGGATTTGTTGTCAGAAACAGGCTAGATTTCTCTAGGAAATCTTAAAGC 2572  
 QY 1199 AAGCAAAA-TTTTGTGCAAGTGAAGCAAACTATCGAAGAAATGCGTTATCAGGC 1257  
 DB 2571 AAGCAAAAATTTTGTGTCAGAGTGAAGCAAACTATCGAAGAAATGCGTTATCAGGC 2512  
 QY 1258 GACCATCGGGCTTTAGAGACTTTTGCAATCATCTTGTCTATGTGAGTTTGGCCTTGA 1317  
 DB 2511 GACCATCGGGCTTTAGAGACTTTTGCAATCATCTTGTCTATGTGAGTTTGGCCTTGA 2452  
 QY 1318 ATGGCAATATGCTTTCAGTCCGATATGGCTTATATCATACCTTTTGGCTACTGTGC 1377  
 DB 2451 ATGGCAATATGCTTTCAGTCCGATATGGCTTATATCATACCTTTTGGCTACTGTGC 2392  
 QY 1378 AGTCTGTTTATAGACATTTCTTTTGAAGAAATCAAAATATGATTTGCAAGCATTTGG 1437  
 DB 2391 AGTCTGTTTATAGACATTTCTTTTGAAGAAATCAAAATATGATTTGCAAGCATTTGG 2332

QY 1438 TGGTTAATGACGTAATGGGGATTCATTAAACAATACCTTGATGATTTGATGCTAT 1497  
 DB 2331 TGGTTAATGACGTAATGGGGATTCATTAAACAATACCTTGATGATTTGATGCTAT 2272  
 QY 1498 TCGTGAAGATCGCCAGCAACCTGTTTACCCCTATGATGATTTTATGATGATCCCT 1557  
 DB 2271 TCGTGAAGATCGCCAGCAACCTGTTTACCCCTATGATGATTTTATGATGATCCCT 2212  
 QY 1558 TCAAAAGACCTTACCGCCAGCGTAATGACACAGCTACATCTATCATGTTTGTAAAT 1617  
 DB 2211 TCAAAAGACCTTACCGCCAGCGTAATGACACAGCTACATCTATCATGTTTGTAAAT 2152  
 QY 1618 GCTTTGTTTATAGGGGCTCTCTGCTTTAATTTTGATTTATATATGACATAGGAT 1677  
 DB 2151 GCTTTGTTTATAGGGGCTCTCTGCTTTAATTTTGATTTATATATGACATAGGAT 2092  
 QY 1678 TCTTCTAGGAATCTTATCGCTCTTATATTTGACACACCTCTGTTTATAGTCCG 1737  
 DB 2091 TCTTCTAGGAATCTTATCGCTCTTATATTTGACACACCTCTGTTTATAGTCCG 2032  
 QY 1738 TAAAGAAATCGCTCAAAATTAAGTACCGTTAACTTAATCTAAAGTGAATATATAA 1797  
 DB 2031 TAAAGAAATCGCTCAAAATTAAGTACCGTTAACTTAATCTAAAGTGAATATATAA 1972  
 QY 1798 ATCTCCTTTGGGACTTATAGCCCAAGGCCCTGTGTATTAATTTATGACAAATTCAG 1857  
 DB 1971 ATCTCCTTTGGGACTTATAGCCCAAGGCCCTGTGTATTAATTTATGACAAATTCAG 1912  
 QY 1858 ATATATGC 1864  
 DB 1911 ATATATGC 1905

RESULT 2  
 AP002547/c  
 LOCUS  
 DEFINITION Chlamydia pneumoniae J138 genomic DNA, complete sequence,  
 section 3/4.  
 ACCESSION AP002547 AB033786 AB033816 AB033817 BA000008  
 VERSION AP002547.2 GI:10176693  
 KEYWORDS  
 SOURCE Chlamydia pneumoniae J138 (strain:J138) DNA.  
 ORGANISM Chlamydia pneumoniae J138  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

REFERENCE  
 AUTHORS  
 1 Shirai, M., Hirakawa, H., Ouchi, K., Tabuchi, M., Kishi, F., Kimoto, M., Takeuchi, A., Nishida, J., Shibata, K., Fujinaga, R., Yoneda, H., Matsushima, H., Tanaka, C., Furukawa, S., Miura, K., Nakazawa, A., Ishii, K., Shiba, T., Hattori, M., Kohara, S. and Nakazawa, T.  
 Comparison of outer membrane protein genes omp and pmp in the whole genome sequences of Chlamydia pneumoniae isolates from Japan and the United States  
 J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)

JOURNAL MEDLINE  
 REFERENCE  
 AUTHORS  
 2 Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., Shiba, T., Ishii, K., Hattori, M., Kohara, S. and Nakazawa, T.  
 Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CML029 from USA  
 Nucleic Acids Res. 28 (12), 2311-2314 (2000)

JOURNAL MEDLINE  
 REFERENCE  
 AUTHORS  
 3 (bases 1 to 300550)  
 Direct Submission  
 Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University School of Medicine, Department of Microbiology, 1-1-1 Minamikogushi, Ube, Yamaguchi 755-8505, Japan  
 (E-mail: mshirai@po.cc.yamaguchi-u.ac.jp, Tel: 81-836-22-2227, Fax: 81-836-22-2415)  
 On or before Sep 15, 2000 this sequence version replaced  
 gi:6172298, gi:6172300, gi:6172396, gi:6172398, gi:8978889.

FEATURES  
 source  
 1. 300550

```

/organism="Chlamydophila pneumoniae J138"
/strain="J138"
/db_xref="taxon:138677"
/notes="synonym:Chlamydia pneumoniae (strain J138)"
complement(21..794)
/gene="dapF"
complement(21..794)
/gene="dapF"
/codon_start=1
/transl_table=11
/product="diaminopimelate epimerase"
/protein_id="BAA98725.1"
/db_xref="GI:8978890"
/translation="MAYFSTSTISKYFIYSGAGNRFLGELTPEVEDVRFLCQTRVD
GFLYKSSCADAGLIIFNSDGSPTWCGNGLRCAIAHLASQKGSKISVSTDSGLYS
GFLYSWRVLVDMTLADRAVSHRESRDPDPKPVVCIHTGVPHAVIILPEISTDL
SILGFLRYHOTSPDGVNWNFVQILGHCOLRUYERYGVEGETAACGTGALASLVV
SNSWGKESIOIHTWGEMLTVSNQRGRVYLQGSVTRDL"
complement(763..1338)
/gene="clpP_1"
complement(763..1338)
/gene="clpP_1"
/codon_start=1
/transl_table=11
/product="Clp protease"
/protein_id="BAA98726.1"
/db_xref="GI:8978891"
/translation="MADDEVHKLRIIEKELLEARRVFFSPVTEKSASDAIKKLWYL
ELKDPGPIVINSVPGSVDAFVNDIOIKMLTSPVTVTVGLAASMGSVLSCAAP
GRRFATPSRIMIHQPSIGGPICTGQATDLIHAREILTKARIIDVVEATNQPRDI
EKALDRMWTANAKAFGLDGLILFNDL"
complement(1358..2917)
/gene="glyA"
complement(1358..2917)
/gene="glyA"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98727.1"
/db_xref="GI:8978892"
/translation="MLKVFKEFKFAIVEITKVAVVUSLLHKLFLNASGKGKOSLAS
TAYLAALDHLNAPSIIGERIIDELKSQRSHLKIASENSSLVSQLMGNLLTDKYC
EGSPKPKFYSCCENDAIEWECVETAKELFADACVQPHSGADANLLAVNAILTHKY
QPAVSKLYGYNVETLNEEYETLLKAENSCVCLGPSLNGGHLTHGNVRLNVMKLM
RCFYDNPDPTECFDYAEISRLAKYEPKVLIAGYSSYRRLNFAVLKQIAEDCGSVL
WDMAHFAGVAGGVFVDEENPIPYADIVTTTHKTLRGGPRGLVLAATREYESTLKA
CPLMMWGPPLPHVIAAKTVALKEALSVDFFKYAHQVNNARELAREFLSHGLRLTGTG
DNHMVVDLGSGLGKIAEDILSSVGIANNLSLPSDAIGKWDTSGLRGTPLATFL
QMGIDEEVADIIKVLNRILRSCHVEGSSKKNGKELPEAIAQARDVRNLLRFP
LYEIDIEALV"
3107..3778
/gene="CPj0522"
3107..3778
/gene="CPj0522"
/codon_start=1
/transl_table=11
/product="CT433 hypothetical protein"
/protein_id="BAA98728.1"
/db_xref="GI:8978893"
/translation="MTLYLGNQNTARKYQAHYLPILTPPYAKSTQNKALQFLQ
ATHVLLSPSTHFLFSNMTSLLSKATLKTICYIGESTKERLLSLFGQKVVYATQ
EIAHGIPPLQALPSSARILYPHSSLARPIRELYNRPFFSYPHYTVPRKLUKNI
LSKYKKIIFTSPTVRAPAKIFRPPKTYWCQGRMTLQEFQKFSQKQVSLLETGK
SRTSP"
3843..4175
/gene="CPj0523"
3843..4175
/gene="CPj0523"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98729.1"
/db_xref="GI:8978894"

```

```

/translation="MASSATPPGPDGTAPSLPPBAPTRPRYNFKLAFVTIAIALWIAL
IATTIAGICIHPLCSFILTALPLYISYICSHYARNVYIALDVVDPHDSKLODMRS
HSPISFSDR"
4223..5302
/gene="CPj0524"
4223..5302
/gene="CPj0524"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98730.1"
/db_xref="GI:8978895"
/translation="MSGPSTESSQVSLSVPRDKELIAPKQFTIAKISTLAILASL
ALGALVAGISLTVLGNPVLALLITLALFVVTVLVIHOMTSKVSNNWOKVLEQNFK
PLAGAKQEKVDCYSNEMQYNNHNPFKVAIOYDASQFPQFTFLGLRVIEKNQST
GIIFNPVPTNLINDTATNLSTILYSLDKGVSDMTCKREGGPAKEDGFPSPTEVRV
VKLPNALDQTNLNLNLSAEKKSILPTFLGHGCPKSEELNQOEYVROALLAVENCL
KAAIESHAALVALPLETSVYVPEEILPKEGIFYWNDQTAQCKRALLDAIQNTALR
YPORSLLVLODPFNIESQSREE"
5642..6406
/gene="CPj0525"
5642..6406
/gene="CPj0525"
/codon_start=1
/transl_table=11
/product="CT398 hypothetical protein"
/protein_id="BAA98731.1"
/db_xref="GI:8978896"
/translation="MHDALLSILAQELDKIMLRVKKEHOKELAKVOSLKSDIRR
KVOEKLEMLNLTQIRDCENRIOEISEQINKLENOQAAVKMDEFNALQEMTANK
ERRSELQSLDMKQAGGEDLVSLKESLASTENSSVIEKIFESIKLINEEGKAL
LEQRTBLKHATNPELLISYIERLLNKKDVRVVFENRVCVSGCHIVLTPQHENVLRKKD
RLIFCEHCSRILYQESQVNAQENSTAKRRRRRAAV"
complement(6827..7816)
/gene="kpsF"
complement(6827..7816)
/gene="kpsF"
/codon_start=1
/transl_table=11
/product="GutQ/kpsF family sugar-P isomerase"
/protein_id="BAA98732.1"
/db_xref="GI:8978897"
/translation="MSPSMISTDVCDILKQKEAVDFPQAFQPKAMQALAEIKLGH
SGVFFSGVSGKCVARKLVATLQSLSERALFFSPVDLLHGLGLVSGPDVCLFSSK
TCQWLEDITVPHLKSRAILVAITSMPSYLNLAALSDIVLPSVAELDPNLIPTNST
TCQWIFDFTLMLLFHSRGSVLSYTKGNHPSQGVKANGKVDKFMFPKTEVPFCHLG
DKYSFSLVPSAYCCCVICVDQFLMGIFTDGLRSLASYGGEVLSLSLEKVMTA
NPRCIEDSDIALALQLMESSSPVAVLPVLDBENRHVTGLLHHTLAKAGLL"
complement(7843..9024)
/gene="sucB_2"
complement(7843..9024)
/gene="sucB_2"
/codon_start=1
/transl_table=11
/product="dihydrolipoamide succinyltransferase"
/protein_id="BAA98733.1"
/db_xref="GI:8978898"
/translation="MIFEPFPKIGETSSGGSIVRWLKLGDHVARDEPLIEVSTDKI
ATLPSPKAGRLVFCVNEGDEVASGDVLGLILEBIESEADDESTCPLTSCETKSEA
SSSSSSWFPSPVLSLAQREISLKSDEPHASLVVDVDTLNLNLSGSRQRFIDTH
IFQGVNRIWSPRLRAAISLKSDEPHASLVVDVDTLNLNLSGSRQRFIDTH
GVKLITTSFTVQLAQTLRQFPLNGSLDGTIVMKKSNVNVAVNLKREGVVPVTH
NCQDRGLVSTAKALADLSSARLNKLPDSEVQSGSVTVTFGMTGALIGMPIRYPEV
AILGIGTIQKRVVVRDDSLAIRKVVVTVLTFDHRVLDGIYVSGEFLTSLKNRLESVTM
G"
complement(9038..10282)
/gene="gltr"
complement(9038..10282)
/gene="gltr"
/codon_start=1
/transl_table=11
/product="glutamate symport"
/protein_id="BAA98734.1"

```

```

/db_xref="GI:8978899"
/translation="MKLMKIRIFGVYTLGLVEDKAIPEKPIGDIPLNLSMNV
PLVFCSMVIGIASISDMKILGRIKSVLYLGTALALVILCEPAMIPEPNDGCPA
QAQMSDAVATVIDSNKTAAYFLJIAQVPSNPVPSFAGNLIQIIFIPIFGLARL
SBERGPVERFIDFSEIMLRWNMIMSAPYGVSNMISGNHGLVLMQDKEFI
AYILCLFATLIVGGIARFGCKMSFKLSMDMDISCAVTSASSALPLPYMBCVS
KMLGSAEVSQVPLPLGATVNNNGRIIPQGMVAVPFAQVNPDLISILLIVATP
SAVGSAGVPGGMITLQSVLASVGLPIQIALILAGIDRLDIVIPMILGDAVVAITI
VASGELSPYESIKQESVEIT"
gene
  complement(10282..11379)
  /gene="ycaH"
  complement(10282..11379)
  /gene="ycaH"
  complement(10282..11379)
  /gene="ycaH"
  /codon_start=1
  /transl_table=11
  /product="Atpase"

Query Match      78.3%; Score 1453; DB 1; Length 300550;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

QY 1 ATGACATTCGCATATGTAGAGAGGGATCAGCGCATGCTAATGACAAATATTCG 60
Db 51024 ATGACATTCGCATATGTAGAGAGGGATCAGCGCATGCTAATGACAAATATTCG 60
QY 61 AAACCGTGATGGCGATGCTGTAGTATGACGCTTAATGCTCAAGCCCTATTTT 120
Db 50964 AAACCGTGATGGCGATGCTGTAGTATGACGCTTAATGCTCAAGCCCTATTTT 120
QY 121 AAACGTCCATGAAAAATATGCGAGTGTCTCAGGAAATTTTCCACCGTGAAGGAG 180
Db 50904 AAACGTCCATGAAAAATATGCGAGTGTCTCAGGAAATTTTCCACCGTGAAGGAG 180
QY 181 CAACTCGCCTCAGATTTAAATCTGAGAGCATGCTTTGTTCCGAGGTTCTCACTGA 240
Db 50844 CAACTCGCCTCAGATTTAAATCTGAGAGCATGCTTTGTTCCGAGGTTCTCACTGA 240
QY 241 AGAGACGATCTCTTCTGATCTTGGAAAAAACAATGACACAGCAATTAATCTCAGATG 300
Db 50784 AGAGACGATCTCTTCTGATCTTGGAAAAAACAATGACACAGCAATTAATCTCAGATG 300
QY 301 CTGTGGCTTGGCAATGCTTAATGTTTGAAGAAGCTAATTAATGATTTGAGGCGTCA 360
Db 50724 CTGTGGCTTGGCAATGCTTAATGTTTGAAGAAGCTAATTAATGATTTGAGGCGTCA 360
QY 361 CGCTTGGGAGCTGTTCTTGAATCTTTGCTTATCTGGGACGCTCAAGATTTTGA 420
Db 50664 CGCTTGGGAGCTGTTCTTGAATCTTTGCTTATCTGGGACGCTCAAGATTTTGA 420
QY 421 TGGCCACATCACTTGTCAAGACTCGCTGGGATGTTCTGTATGGGATGCGCGTGA 480
Db 50604 TGGCCACATCACTTGTCAAGACTCGCTGGGATGTTCTGTATGGGATGCGCGTGA 480
QY 481 TGCATAATGTTTGTATTCGAAGAATCCGAGAGAAATTTTATGTCGCTTAA 540
Db 50544 TGCATAATGTTTGTATTCGAAGAATCCGAGAGAAATTTTATGTCGCTTAA 540
QY 541 AAAATCTGTAGAAAAAGATATACCAAGCTTTTGGAGCCATTTTGAATCTTAAGTAC 600
Db 50484 AAAATCTGTAGAAAAAGATATACCAAGCTTTTGGAGCCATTTTGAATCTTAAGTAC 600
QY 601 TACAGTATGGCTCAGCAGCTTCTTCTCTAGATACAGGCGCTATTAAGGGTTTGC 660
Db 50424 TACAGTATGGCTCAGCAGCTTCTTCTCTAGATACAGGCGCTATTAAGGGTTTGC 660
QY 661 TTGACATGATTTTGAAGATTTTCTCTCAATGTTTACGGCTCTTTTCAAGACTTAA 720
Db 50364 TTGACATGATTTTGAAGATTTTCTCTCAATGTTTACGGCTCTTTTCAAGACTTAA 720
QY 721 TTGTCATGCTGTGATGATTAAGCAACCATACATCAGTGTGATGATGAATTAAGT 780
Db 50304 TTGTCATGCTGTGATGATTAAGCAACCATACATCAGTGTGATGATGAATTAAGT 780

```

```

QY 781 CGGGGGAATAAGCATGATTTCTGTAGAGGATGACAAAAAATTGGGCTGTTCTGGAAG 840
Db 50244 CGGGGGAATAAGCATGATTTCTGTAGAGGATGACAAAAAATTGGGCTGTTCTGGAAG 840
QY 841 TGTGTTTCTTTTATAGTTGGCTGCTCTCGGGTTTGGAGCTGGAATTCGGTTTGGGAAT 900
Db 50184 TGTGTTTCTTTTATAGTTGGCTGCTCTCGGGTTTGGAGCTGGAATTCGGTTTGGGAAT 900
QY 901 GGAATTTAAAGAGGGATGCTTACCTTTAATCCAAAGAGCATGAGATCAGGATGT 960
Db 50124 GGAATTTAAAGAGGGATGCTTACCTTTAATCCAAAGAGCATGAGATCAGGATGT 960
QY 961 TGCTCAATGCGGCAAGTGTGATTAATCACTACAGAGAGCGGCTTCTCTGAGA 1020
Db 50064 TGCTCAATGCGGCAAGTGTGATTAATCACTACAGAGAGCGGCTTCTCTCTGAGA 1020
QY 1021 CTTCGCTATTCAAACATTTTGAATCTTCAGAAAAAGATCAAAATCTATTTTATGATTAAC 1080
Db 50004 CTTCGCTATTCAAACATTTTGAATCTTCAGAAAAAGATCAAAATCTATTTTATGATTAAC 1080
QY 1081 TTTAAGCTTACT-AGCGATACAGAGCTTCTCTCTTAAATTAACATCAGAGCTGCG 1139
Db 49944 TTTAAGCTTACTTAAAGAGATACAGAGCTTCTCTCTTAAATTAACATCAGAGCTGCG 1139
QY 1140 GTTA-TTGGGAGTGTGTCAAGAAACAGGCTAGATTTCTACGGAACCTTAAACGA 1198
Db 49884 GTTA-TTGGGAGTGTGTCAAGAAACAGGCTAGATTTCTACGGAACCTTAAACGA 1198
QY 1199 AACGCAAAA-TTTTGTCAAGGTAAAGCAAGCAAACTATGCAAGAAATCGTTATCAGCG 1257
Db 49824 AACGCAAAA-TTTTGTCAAGGTAAAGCAAGCAAACTATGCAAGAAATCGTTATCAGCG 1257
QY 1258 GACCATCGGCTTTTGAAGCTTTGGCAATCATCTTGGCTTATGATGATTTGGCTTTGA 1317
Db 49764 GACCATCGGCTTTTGAAGCTTTGGCAATCATCTTGGCTTATGATGATTTGGCTTTGA 1317
QY 1318 ATGGCAATATGCTTACAGTGGCGTATGCGCTTAATTCATGACCTTTGGCTACCTGTCG 1377
Db 49704 ATGGCAATATGCTTACAGTGGCGTATGCGCTTAATTCATGACCTTTGGCTACCTGTCG 1377
QY 1378 AGCTTGTATTAAGCAATTTCTTTTGAAGAAATTCAAATGATTTGCAAGCCATTGG 1437
Db 49644 AGCTTGTATTAAGCAATTTCTTTTGAAGAAATTCAAATGATTTGCAAGCCATTGG 1437
QY 1438 TGCCTTAATGACTTATTTGGGGTATTCATTAACAATACTTTGATGATTTTGAATGCTAT 1497
Db 49584 TGCCTTAATGACTTATTTGGGGTATTCATTAACAATACTTTGATGATTTTGAATGCTAT 1497
QY 1498 TCGTGAAGATCGCAAGCGCAAGCTGTTTACCCCTATGCAATGTTTATGATGATGCGCT 1557
Db 49524 TCGTGAAGATCGCAAGCGCAAGCTGTTTACCCCTATGCAATGTTTATGATGATGCGCT 1557
QY 1558 TCAAAAGACGTTACGCGCAGCGTAATGACACAGCTACCACTATCAAGTTTGTAAAT 1617
Db 49464 TCAAAAGACGTTACGCGCAGCGTAATGACACAGCTACCACTATCAAGTTTGTAAAT 1617
QY 1618 GCTTTGTTTATAGGGGCTCTCTGCTCTTAAATTTTGCATTTATATGACCAATAGGAT 1677
Db 49404 GCTTTGTTTATAGGGGCTCTCTGCTCTTAAATTTTGCATTTATATGACCAATAGGAT 1677
QY 1678 TCTTCTAGGAACCTTATGCTCTCTTATATATGACCAACCTGTTGTTTATGATGCG 1737
Db 49344 TCTTCTAGGAACCTTATGCTCTCTTATATATGACCAACCTGTTGTTTATGATGCG 1737
QY 1738 TAAAGAAATCGCTCAAAATTAAGTACCGTTAACTTAATCTAACGTTAGCAATATATAA 1797
Db 49284 TAAAGAAATCGCTCAAAATTAAGTACCGTTAACTTAATCTAACGTTAGCAATATATAA 1797
QY 1798 ATCTCTTTGGACCTTATGATCCCAAGGCGCTGTGTATTAATTTATGACAAATTCAG 1857
Db 49224 ATCTCTTTGGACCTTATGATCCCAAGGCGCTGTGTATTAATTTATGACAAATTCAG 1857
QY 1858 ATATATGC 1864

```





```

/db xref="GI:7189113"
/translacion: "MTYALINDPVDSLATNNABSKPSPLOLPNHVALIMGNRWY
KAREECGHTHSGHYGAKVLPNIIANVLDLITKVLTLTSTENPGPKKEIJOEIT
NIFVTDQLQPLYMENELCLRCIGDLSKLPKGIQKINHVSMSTASFSRLBLIVAVN
YGGKDLVRAFKLHYDILNKISIDSLSDLSLSDLSGLDPLLIARTGEMKMS
NPLMOLATVELYITDMLMPDFTPODLFEALINVQGRSRGK"
complement (6214, .6363)
/gene="CP0184"
/complement (6214, .6363)
/gene="CP0184"
/complete (6214, .6363)
/notes="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AF38057.1"
/db xref="GI:7189114"
/translacion="MPKKPLVSFEVRLKNFLLEPPICLENRIQPLPTGAGEKSFVI
FVLGK"
6637..7737
/gene="CP0185"
/complete (6637..7737)
/gene="CP0185"
/notes="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AF38058.1"
/db xref="GI:7189115"
/translacion="MKPRRISIDLIPTQVITWVRGGNHYVPAONLPRKILGVL
ACGLALISCAFAAGVCTTPFCIGMIGLVLPGLAYIQYKMSRPRPRFRRTX
VPEKPINMLGCLSLQSWKIRPGCYHPCPEVEICSGQELVITIKFQKSRNRSI
FLIENDQALRQIEKSLSRKTPAIDPSVSLSEIQRQOYLDPKVISWSSD
QASDRTHPKSAIYVNIISDAQEPQRCYIDAATYKAFVTDQIDPNIVKHTIYVLT
PLIGVDPALPKRQENILKLSQAFLYSARQVAKRMRBEKQDSIRIKFTITDPTSPS
LYSPHHSRPHSVTPISLSGFVGEDESYFA"
7917..13125
/gene="CP0186"
/complete (7917..13125)
/notes="similar to GP:3220156; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="secDF protein, putative"
/protein_id="AF38059.1"
/db xref="GI:7189116"
/translacion="MKOKVKNRAIITICVPAALAYVLTPTCLYAKRDLKIDKNEAE
HIIKSPKQAOVRKDIVPRVSAIISLHLRGTIQQHPAIPDIVSRFKGEDAEDPT
GNLVGHPNPIKISARLHVGVYREHDDHVIQVASSINTLSVDSFVYSSENEOF
MASSILORVVSACTPCKDCSCSPSIWETAPKEQLQYAKMLSGFEVFSRLAF
COOSFSSNDRLAFLSRLSISNDADIVDOKLKSIVTTLSCITACIRLDCPYEG
LRDCSSSLFSPSIEYCPKREKIFLILHSDILAORTSKRDLRDPDSLAVEKOL
SKULTVQVDEYNNRFSFQMDKDTQGLITIOGERLIGIAEHLTALITHRPAESDL
IPENFVETCRPSESAFVGCYIFSPNTDCNFSKGSYITLLKGRSIVAKYQGGK
LOSEFKDLQNLNCFSTHETAIISWLGDOVLEIRHPLOQLDVGSEFVYKSGCAFL
EVKDIQDRIATVNOIEKRNOSDLVRMHEQYRHKCSMDLOERISAPIPYONLEFNMK
LNMKFRGENIRLIGIDFVGRQLSLFSDHOKGLTDKEDILTKVDELCALINKG
VSEIELRRNDYIHLVSPTSISSEILTSKMSFHVNERSSVSASNYEVQFRLD

```

Query Match 76.2%; Score 1421; DB 1; Length 12198;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1831; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

```

QY 1 ATGACCTCCGCAATATCTCGAGAGGGAGTACGGGCACTGCTATGACATATTCGCG 60
DB 10364 ATGACCTCCGCAATATCTCGAGAGGGAGTACGGGCACTGCTATGACATATTCGCG 10423
QY 61 AAACCGTGAATGCGGTATGCTGTAGTATGACGTTATATGCTACAGACCCCTATTTT 120
DB 10424 AAACCGTGAATGCGGTATGCTGTAGTATGACGTTATATGCTACAGACCCCTATTTT 10483

```

```

QY 121 AAACGTCCTATGAAAAATCATGCCAGTGTCTCAGGAAATTTACCACCGTGAAGTAG 180
DB 10484 AAACGTCCTATGAAAAATCATGCCAGTGTCTCAGGAAATTTACCACCGTGAAGTAG 10543
QY 181 CAACTGCGCTCAGATTTTAAATCTGAGACGATGTCTTTTGTCCGAGGTTCTCAGTGA 240
DB 10544 CAACTGCGCTCAGATTTTAAATCTGAGACGATGTCTTTTGTCCGAGGTTCTCAGTGA 10603
QY 241 AGAGACGATCTCTTCTGATCTGGGAAAAAACAATGTACACAGGATATCTCAGCATG 300
DB 10604 AGAGACGATCTCTTCTGATCTGGGAAAAAACAATGTACACAGGATATCTCAGCATG 10663
QY 301 CTGTGCTGGCAATGCTTATGTTTGAAGAGCTATATTAATTTAGATTGAGAGCGCTCAT 360
DB 10664 CTGTGCTGGCAATGCTTATGTTTGAAGAGCTATATTAATTTAGATTGAGAGCGCTCAT 10723
QY 361 CGCTTGAGAGCTGTCTTCTGATCTTTTGTCTTATCTGGGACGCTTACAGTATTTGGA 420
DB 10724 CGCTTGAGAGCTGTCTTCTGATCTTTTGTCTTATCTGGGACGCTTACAGTATTTGGA 10783
QY 421 TCGCCCACTCAGCTTGTCAAGACTCGCTGGAGATTTGTTCTGTATGGGATGGCCGTAGA 480
DB 10784 TCGCCCACTCAGCTTGTCAAGACTCGCTGGAGATTTGTTCTGTATGGGATGGCCGTAGA 10843
QY 481 TGCAAATGTTCTGTATGAAAGAAATCCGAGAGAAATTTTATGTCTCAAACTCTTA 540
DB 10844 TGCAAATGTTCTGTATGAAAGAAATCCGAGAGAAATTTTATGTCTCAAACTCTTA 10903
QY 541 AAATCTGTAGAAAAAGATATACCAAGCTTTTGGAGCATTTTGTATCTTAATCTTAC 600
DB 10904 AAATCTGTAGAAAAAGATATACCAAGCTTTTGGAGCATTTTGTATCTTAATCTTAC 10963
QY 601 TACAGATTTGGCTCGACACTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 660
DB 10964 TACAGATTTGGCTCGACACTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 11023
QY 661 TTTGACATGATTTTGAATTTCTCTCAATGTTTGAAGGCTTTTCAATGATCAATTA 720
DB 11024 TTTGACATGATTTTGAATTTCTCTCAATGTTTGAAGGCTTTTCAATGATCAATTA 11083
QY 721 TTTCTTCATGCTGTGATGATTAAGACCAACATACAGATTGATGATTAAGTT 780
DB 11084 TTTCTTCATGCTGTGATGATTAAGACCAACATACAGATTGATGATTAAGTT 11143
QY 781 CGTGGGATTAAGCATGATTTCTTGAAGAGTCAAAAACTTTGGCGCTTTCTTGAAG 840
DB 11144 CGTGGGATTAAGCATGATTTCTTGAAGAGTCAAAAACTTTGGCGCTTTCTTGAAG 11203
QY 841 TGTCTTTCTTTTGTGTTGCTGCTCGGGGTTTGGAGCTGTGAATTCGTTTGGGAAT 900
DB 11204 TGTCTTTCTTTTGTGTTGCTGCTCGGGGTTTGGAGCTGTGAATTCGTTTGGGAAT 11263
QY 901 GGAATTTTAAAGAGGATGCTCTTAACTTTAATCCAAAAAGACATGACATGCGCATGT 960
DB 11264 GGAATTTTAAAGAGGATGCTCTTAACTTTAATCCAAAAAGACATGACATGCGCATGT 11323
QY 961 TGTCTCAATCGCTGGCAAGTGTGTGATTAACACGAAAGCTGTCTTTCTTGAAGA 1020
DB 11324 TGTCTCAATCGCTGGCAAGTGTGTGATTAACACGAAAGCTGTCTTTCTTGAAGA 11383
QY 1021 CTTCGCAATCAAACTTTGGATCTTCAGAAAAGATCAAAATCTATTTAGATGAAG 1080
DB 11384 CTTCGCAATCAAACTTTGGATCTTCAGAAAAGATCAAAATCTATTTAGATGAAG 11443
QY 1081 TTTAAGCTATACT-AAGCAGATACGAGCTCTCTCTTAATTAATTAAGATCATGAGCTGCG 1139
DB 11444 TTTAAGCTATACTAAGCAGATACGAGCTCTCTCTTAATTAATTAAGATCATGAGCTGCG 11503
QY 1140 GTTA-TTGTGGATGTGTGTCAAAAACAGGCTTGAATTTCTTACGAAAACCTTAAACGA 1198
DB 11504 GTTA-TTGTGGATGTGTGTCAAAAACAGGCTTGAATTTCTTACGAAAACCTTAAACGA 11563
QY 1199 AAGGCAAAA-TTTTGTGCAAGGTAAGCAACATATGCAAAAATGTGCTTATACAGC 1257

```





/product="glutamyl-tRNA synthetase"  
/protein\_id="AAF38063.1"  
/db\_xref="GI:7189121"  
/translation="MMENVRYRVAAPTGDHPVGTAYMALFNEIFAKRFGKGMILRI  
EDTRSRHODYEENIFSAIRMGCIQMDGDPVGGPYGYSROSEKTIQGVETLLK  
TPCAKCFATPOELAMBAVASTLGRGGDYRRLYSDEEVAAREAAQOYITRLKY  
PLSGCVPRHYSGKGRVFPVDDOVLKSGFPRHYRANVYIDHLMKITHLNGEE  
WLSSTPKHLLIYARGMPEPVLHMFLENPDGTAKSKRANPTSIIFYYDSCYKGA  
VNFPLTMGSMGDEEVSLERTIETENPRRIQKSGAVDIOQLDMNKHYNHESRP  
ECLIKELQGMNLNDEFFELKILPCQSRITTLAEFINLTSFFSGLEYVEBLDQAL  
SEKAIKILSYVKYLEKTDQWTKETCYGSKWLAQAFVNHKKAILPILYVAITGKK  
QGLPLFDSYEIKKPRARALVYAEKLGVPKLAATVDKFMQREDFEAFEDL"  
/complement(6199..6387)  
/gene="CP0191"  
/complement(6199..6387)  
/note="conserved hypothetical protein; identified by  
Glimmer2: putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF38064.1"  
/db\_xref="GI:7189122"  
/translation="MKLFYFCNIIPLTSLPSVSCHTPPPRSYLLAQGRPLVNXKAS  
LEPFLARSFNTCYNTNL"  
/gene="CP0192"  
/complement(6645..6752)  
/note="hypothetical protein; identified by Glimmer2:  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF38065.1"  
/db\_xref="GI:7189123"  
/translation="MFSKSKSFNTKTIYLEKTIKIFICNIYHGVLY"  
/gene="CP0193"  
/complement(6829..7101)  
/note="similar to PTD:1783381; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="cysteine rich outer membrane protein 3"  
/protein\_id="AAF38066.1"  
/db\_xref="GI:7189124"  
/translation="MKAVYLIAAMFGGVVSLSSCCRIYVCCPEPCAPSSCNCEVIR  
KERSCGNACGVSIPSCNPGSTECNSQSPVQKCTSPDKCKQ"  
/gene="CP0194"  
/complement(7165..7281)  
/note="hypothetical protein; identified by Glimmer2:  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF38067.1"  
/db\_xref="GI:7189125"  
/translation="MSLDIGISYRMFIFPFCRVECFVNSYPOHLGLSELRL"  
/gene="CP0195"  
/complement(7268..8538)  
/note="similar to SP:P23700 PTD:550566 PTD:1326243  
PID:1326245 GB:AE001363; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="60 kDa outer membrane protein"  
/protein\_id="AAF38068.1"

/db\_xref="GI:7189126"  
/translation="MSKILRRVTVLALTSMAQSGRIEAAVESLITKIASET  
KPAVPTAKVRLVRNRKQVEQKSCAFQDKERYPEEERQDPVEAQSSCTGRLY  
SVKVNDDCNVEIQSVPEYATVGSPPYIEILAIQKDVDAVITIQLECEAFVSSDP  
ETTPSDKDLWKIDRLGAGDKCKITVWKPDKRECCPTAAVTCAPBLRKYKGGP  
AICIKQEPDCACLRCPCVYKIEVNTGSAIARVTVNPPVDGYSAGORVLSFNL  
GDMRGDKKVFVEFCPPORGOITNVAVTVCGGKGSANVTVNECCOVNINSAD  
MSYCKPVEYSISVNPGLDVLHVIODTLPSGVTVLEAPGGEICCKKVMRIKEMC  
PGETIQFLVYKAQVPGPFTNOVAVTSNCGTCSCAETTHKGLAATMKCULDITN  
DPTICGVETVIRICVTNNGSADNVNLSILKFSKELQPIASSGPTKGTISGTVFEDA  
LPKIKSKSVFSEVYTLKGIAPGDARGAIISSDILTSVSDENTHYV"  
/gene="CP0196"  
/complement(9217..9807)  
/note="similar to SP:P18587; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="15 kDa cysteine-rich outer membrane protein"  
/protein\_id="AAF38069.1"  
/db\_xref="GI:7189127"  
/translation="MSSNMLHPVGGTGTGAABESVNIIVEBIAASGVYAGIQATSS  
PGVMVLIGMAKTKFIQPIRESKLFQSRACQITLVGLILVAGLACMFIPIHSDGA  
NAFWLITPAIIGLILVLTSLCPDRACTSEKLMWQKAGVLEQDLDDGILNNSKIF  
GHVKTENTSRATTPVNDGRTIPVLSVSKIARY"  
/complement(9885..11831)  
gene  
Query Match  
Best Local Similarity 4.9%; Score 92; DB 1; Length 11938;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1773 TAATCAAGCTGAGCAATATAAAATTCCTTGCGACTTATGTCCAAAGCCCTGT 1832  
Db 1 TAATCAAGCTGAGCAATATAAAATTCCTTGCGACTTATGTCCAAAGCCCTGT 60  
Qy 1833 GGTATTAATTTATGCAAAATCAGATAATGC 1864  
Db 61 GGTATTAATTTATGCAAAATCAGATAATGC 92  
RESULT 5  
AF019977/c 3183 bp mRNA linear VMT 15-NOV-1997  
LOCUS  
DEFINITION Gallus gallus smoothened mRNA, partial cds.  
ACCESSION AF019977  
VERSION AF019977.1 GI:2618974  
KEYWORDS  
SOURCE  
ORGANISM Gallus gallus.  
Gallus gallus.  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 3183)  
REFERENCE  
Quirk,J., van den Heuvel,M., Henrique,D., Mariño,V., Sheer,D.,  
Tabin,C. and Ingham,P.W.: The smoothened gene and Hedgehog Signal Transduction in Drosophila  
and Vertebrate Development Cold Spring Harb. Symp. Quant. Biol. (1997) In press  
2 (bases 1 to 3183)  
AUTHORS Quirk,J., van den Heuvel,M., Henrique,D., Mariño,V., Sheer,D.,  
Tabin,C. and Ingham,P.W.  
TITLE Direct Submision  
JOURNAL Submitted (18-AUG-1997) Molecular Embryology Laboratory, Imperial  
REFERENCE Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2 3PX,  
United Kingdom  
COMMENT On Nov 15, 1997 this sequence version replaced gi:2582358.  
FEATURES  
source  
CDs  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
/note="putative trans membrane protein; similar to



```

repeat_region      /rpt_family="MER66A-int"
complement(7740. .8116)
/rpt_family="THE1C"
repeat_region      complement(8117. .8450)
/rpt_family="MER66A-int"
complement(8533. .8821)
/rpt_family="MER66A-int"
repeat_region      complement(8624. .9097)
/rpt_family="MER66A-int"
repeat_region      complement(9273. .9465)
/rpt_family="MER66B-int"
complement(9482. .9919)
/rpt_family="MER66B-int"
repeat_region      complement(9928. .10257)
/rpt_family="MER66B"
repeat_region      complement(10385. .10634)
/rpt_family="HAL1b"
repeat_region      complement(11492. .12021)
/rpt_family="L1M4"
repeat_region      complement(12057. .12098)
/rpt_family="AT_rich"
repeat_region      complement(12123. .12199)
/rpt_family="MER5B"
repeat_region      complement(12247. .13372)
/rpt_family="L1MA4"
repeat_region      complement(13723. .13927)
/rpt_family="LTR27"
repeat_region      complement(13928. .14234)
/rpt_family="A1y"
repeat_region      complement(14235. .14575)
/rpt_family="LTR27"
repeat_region      complement(14608. .14754)
/rpt_family="LTR27"
repeat_region      complement(14757. .14825)
/rpt_family="L1MA4"
repeat_region      complement(14832. .14879)
/rpt_family="AT_rich"
repeat_region      complement(15041. .15400)
/rpt_family="L1ME4A"
repeat_region      complement(15674. .15748)
/rpt_family="MIR"
repeat_region      complement(17417. .17437)
/rpt_family="AT_rich"
repeat_region      complement(17759. .18249)
/rpt_family="LOR1a"
repeat_region      complement(18656. .18809)
/rpt_family="FRAM"
repeat_region      complement(18878. .18905)
/rpt_family="(T)n"
repeat_region      complement(19369. .19479)
/rpt_family="L1PB3"
repeat_region      complement(19841. .20171)
/rpt_family="MER21B"
repeat_region      complement(20172. .20498)
/rpt_family="Alusx"
repeat_region      complement(20499. .20538)
/rpt_family="MER21B"
repeat_region      complement(20539. .21022)
/rpt_family="MER66B"
repeat_region      complement(21023. .21485)
/rpt_family="MER21B"
repeat_region      complement(22732. .22820)
/rpt_family="L2"
repeat_region      complement(22990. .23052)
/rpt_family="L2"
repeat_region      complement(24042. .24067)
/rpt_family="AT_rich"
repeat_region      complement(24261. .24921)
/rpt_family="MER21B"
repeat_region      complement(26131. .26365)
/rpt_family="MIR"
repeat_region      complement(26415. .26475)
/rpt_family="MIR"

Query Match      1.2%: Score 22; DB 9; Length 151032;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1395 ATTCTTTTGAAGAAATTCA 1416
Db 19872 ATTCTTTTGAAGAAATTCA 19851

RESULT 7
AC124363/c      159198 bp      DNA      linear      HTG 20-JUN-2002
LOCUS      Mus musculus chromosome UNK clone RP24-531A9, WORKING DRAFT
DEFINITION      SEQUENCE, 1 unordered piece.
ACCESSION      AC124363
VERSION      AC124363.1 GI:21426484
KEYWORDS      HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULFILLTOP.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 159198)
AUTHORS      McPherson,J.D. and Waterston,R.H.
TITLE      The sequence of Mus musculus clone
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 159198)
AUTHORS      McPherson,J.D. and Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE      3 (bases 1 to 159198)
AUTHORS      McPherson,J.D. and Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BB0531A09
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 158386 bases at least Q40
Consensus quality: 158639 bases at least Q30
Consensus quality: 158759 bases at least Q20
Insert size: 13800; agarose-fp
Insert size: 159198; sum-of-ctnigs
Quality coverage: 15.36 in Q20 bases; agarose-fp
Quality coverage: 13.93 in Q20 bases; sum-of-ctnigs

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. the true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 159198: contig of 159198 bp in length.

#### FEATURES

source  
Location/Qualifiers  
1. 159198  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="XNK"  
/clones="RP24-531A9"  
misc\_feature  
1. 159198  
/note="assembly name:Contig8"

BASE COUNT 44143 a 34841 c 35294 g 44919 t 1 others  
ORIGIN

Query Match 1.2%; Score 22; DB 2; Length 159198;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 831 TTTCTGGAAGTCTTTTCTTTT 852  
|||||  
Db 39433 TTTCTGGAAGTCTTTTCTTTT 39412

RESULT 8  
AC036125 172579 bp DNA linear PRI 24-AUG-2002  
LOCUS  
DEFINITION Homo sapiens chromosome 18, clone RP11-482F20, complete sequence.  
AC036125  
VERSION  
AC036125.7 GI:17223368  
KEYWORDS  
HTG.  
SOURCE  
human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 172579)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens chromosome 18, clone RP11-482F20  
Unpublished  
2 (bases 1 to 172579)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F., Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N., Plean, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 172579)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,

#### TITLE

JOURNAL

REFERENCE

AUTHORS

Direct Submission  
Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 172579)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

#### TITLE

JOURNAL

COMMENT

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 1, 2001 this sequence version replaced gi:16949121.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence submissions@genome.wi.mit.edu

Center project name: L7322

Center clone name: 482\_F\_20

#### FEATURES

source

Location/Qualifiers  
1. 172579  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"

/clone="RP11-482F20"

/clone lib="RP11-11 Human Male BAC"

1205\_1667

/rpt\_family="MER21B"

1668\_2151

/rpt\_family="MER66B"

repeat\_region

repeat\_region

```

repeat_region      2152..2191
/rpt_family="MER21B"
complement(2192..2517)
repeat_region      /rpt_family="AluSx"
2518..2848
/rpt_family="MER21B"
repeat_region      3210..3320
/rpt_family="L1PB3"
3784..3812
/rpt_family="(A)n"
repeat_region      3881..4034
/rpt_family="FRAM"
4441..4931
/rpt_family="LOR1A"
5253..5273
/rpt_family="AT_rich"
repeat_region      complement(6942..7016)
/rpt_family="MIR"
complement(7290..7649)
/rpt_family="L1ME4A"
7811..7858
/rpt_family="AT_rich"
complement(7865..7933)
/rpt_family="L1MA4"
repeat_region      complement(7936..8082)
/rpt_family="LTR27"
complement(8115..8455)
/rpt_family="LTR27"
8456..8762
/rpt_family="AluY"
complement(8763..8967)
/rpt_family="LTR27"
repeat_region      complement(8968..10443)
/rpt_family="L1MA4"
10491..10567
/rpt_family="MER5B"
10592..10633
/rpt_family="AT_rich"
complement(10669..11198)
/rpt_family="L1M4"
repeat_region      complement(12056..12305)
/rpt_family="HAL1b"
12433..12762
/rpt_family="MER6B"
12771..13208
/rpt_family="MER6B-int"
13325..13417
/rpt_family="MER6A-int"
13593..13866
/rpt_family="MER6A-int"
13869..14157
/rpt_family="MER6A-int"
14240..14573
/rpt_family="MER6A-int"
14574..14950
/rpt_family="THE1C"
14951..15490
/rpt_family="MER6A-int"
15491..15576
/rpt_family="MER6A-int"
15577..15807
/rpt_family="MER6A-int"
16025..16375
/rpt_family="MER6A-int"
16413..16824
/rpt_family="MER6A"
16966..16994
/rpt_family="AT_rich"
complement(17062..17359)
/rpt_family="AluSg"
17670..17692
/rpt_family="AT_rich"
17719..17753

```

---

```

repeat_region      /rpt_family="(CA)n"
complement(18579..18870)
/rpt_family="AluSx"
repeat_region      20083..20107
/rpt_family="(CA)n"
repeat_region      21436..21613
/rpt_family="L1MD2"
repeat_region      22014..22159
/rpt_family="(TA)n"
repeat_region      complement(22160..22441)
/rpt_family="AluSc"
22776..23232
repeat_region

```

```

Query Match      1.2%: Score 22; DB 9; Length 172579;
Best Local Similarity 100.0%: Pred. No. 2.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1395 ATTCTTTTGAAGAAATTC A 1416
Db      2817 ATTCTTTTGAAGAAATTC A 2838

```

```

RESULT 9
AC121084
LOCUS
DEFINITION
Mus musculus clone RP23-191E14, WORKING DRAFT SEQUENCE, 21 ordered
pieces.
AC121084
AC121084.3 GI:20986564
VERSION
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
Mus musculus.
SOURCE
Mus musculus.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
1 (bases 1 to 209636)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-191E14
Unpublished
2 (bases 1 to 209636)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barina,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhagalter,B., Brown,A., Camarata,J., Campiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,
Lander,T., Lehoczeky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Menus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Riese,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Strausse,N., Subramaniam,A., Talams,J., Testfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggilio,U., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 209636)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barina,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhagalter,B., Brown,A., Camarata,J., Campiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,
Lander,T., Lehoczeky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Menus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Riese,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Strausse,N., Subramaniam,A., Talams,J., Testfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggilio,U., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

```

Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K., Lanazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneses, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vasiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, W.

Direct Submission  
Submitted (20-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 20, 2002 this sequence version replaced gi:20901891.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

# TITLE JOURNAL COMMENT

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIGR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submission@genome.wi.mit.edu](mailto:sequence_submission@genome.wi.mit.edu)  
----- Project Information  
Center project name: L21012  
Center clone name: 191.E.14

----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 202259 bases at least Q40  
Consensus quality: 205316 bases at least Q30  
Consensus quality: 206630 bases at least Q20  
Insert size: 207636; sum-of-contigs  
Quality coverage: 7.6 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 21 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

\* 1 12959: contig of 12959 bp in length  
\* 12960 13059: gap of 100 bp  
\* 13060 13748: contig of 689 bp in length  
\* 13749 13848: gap of 100 bp  
\* 13849 14975: contig of 1127 bp in length  
\* 14976 15075: gap of 100 bp  
\* 15076 16142: contig of 1067 bp in length  
\* 16143 16242: gap of 100 bp  
\* 16243 17810: contig of 1568 bp in length  
\* 17811 17910: gap of 100 bp  
\* 17911 18957: contig of 1047 bp in length  
\* 18958 19057: gap of 100 bp  
\* 19058 20469: contig of 1412 bp in length  
\* 20470 20569: gap of 100 bp  
\* 20570 22600: contig of 2031 bp in length  
\* 22601 22700: gap of 100 bp  
\* 22701 24776: contig of 2076 bp in length  
\* 24777 24876: gap of 100 bp  
\* 24877 29447: contig of 4571 bp in length  
\* 29448 29547: gap of 100 bp  
\* 29548 38261: contig of 8714 bp in length  
\* 38262 38361: gap of 100 bp  
\* 38362 46977: contig of 8616 bp in length  
\* 46978 47077: gap of 100 bp  
\* 47078 57145: contig of 10068 bp in length

\* 57146 57245: gap of 100 bp  
\* 57246 68889: contig of 11644 bp in length  
\* 68890 68989: gap of 100 bp  
\* 68990 82201: contig of 13212 bp in length  
\* 82202 82301: gap of 100 bp  
\* 82302 97077: contig of 14776 bp in length  
\* 97078 97177: gap of 100 bp  
\* 97178 109416: contig of 12239 bp in length  
\* 109417 109516: gap of 100 bp  
\* 109517 124991: contig of 15475 bp in length  
\* 124992 125091: gap of 100 bp  
\* 125092 151033: contig of 25942 bp in length  
\* 151034 151333: gap of 100 bp  
\* 151334 185034: contig of 33901 bp in length  
\* 185035 185134: gap of 100 bp  
\* 185135 209636: contig of 24502 bp in length.

## FEATURES

source  
1. .209636  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-191E14"  
/clone\_lib="RPC1-23 Female Mouse BAC"  
misc\_feature  
1. .12959  
/note="assembly\_fragment  
vector\_side:left"  
13060. .13748  
/note="assembly\_fragment"  
13849. .14975  
/note="assembly\_fragment"  
15076. .16142  
/note="assembly\_fragment"  
16243. .17810  
/note="assembly\_fragment"  
17911. .18957  
/note="assembly\_fragment"  
19058. .20469  
/note="assembly\_fragment"  
20570. .22600  
/note="assembly\_fragment"  
22701. .24776  
/note="assembly\_fragment"  
24877. .29447  
/note="assembly\_fragment"  
29548. .38261  
/note="assembly\_fragment"  
38362. .46977  
/note="assembly\_fragment"  
47078. .57145  
/note="assembly\_fragment"  
57246. .68889  
/note="assembly\_fragment"  
68990. .82201  
/note="assembly\_fragment"  
82302. .97077  
/note="assembly\_fragment"  
97178. .109416  
/note="assembly\_fragment"  
109517. .124991  
/note="assembly\_fragment"  
125092. .151033  
/note="assembly\_fragment"  
151134. .185034  
/note="assembly\_fragment"  
185135. .209636  
/note="assembly\_fragment  
vector\_side:right"  
clone\_end:T7

BASE COUNT 58935 a 45360 c 45624 g 57716 t 2001 others  
ORIGIN

Query Match 1.2%; Score 22; DB 2; Length 209636;  
Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 831 TTTCTGGAAGTGTTCCTTT 852  
Db 88632 TTTCTGGAAGTGTTCCTTT 88653

RESULT 10  
AC121990 247074 bp DNA linear HTG 21-MAY-2002  
LOCUS Mus musculus chromosome UNK clone RP24-298L8, WORKING DRAFT  
DEFINITION  
AC121990  
SEQUENCE, 15 unordered pieces.  
AC121990  
AC121990.1 GI:21040114  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS  
Mus musculus.  
SOURCE  
ORGANISM  
Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 247074)  
McPherson, J.D. and Waterston, R.H.  
The sequence of Mus musculus clone  
Unpublished  
2 (bases 1 to 247074)  
McPherson, J.D. and Waterston, R.H.  
Direct Submission  
Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

COMMENT  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
Contact: submissions@wustl.wustl.edu  
Project Information  
Center project name: M\_BB0298L08  
----- Summary Statistics -----  
Sequencing vector: M13; 0%  
Sequencing vector: plasmid; 100%  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.99019  
Consensus quality: 240193 bases at least Q40  
Consensus quality: 241582 bases at least Q30  
Consensus quality: 242647 bases at least Q20  
Insert size: 18300; agarose-fp  
Insert size: 245674; sum-of-contigs  
Quality coverage: 9.96 in Q20 bases; agarose-fp  
Quality coverage: 6.54 in Q20 bases; sum-of-contigs  
----- NOTE: This is a 'working draft' sequence. It currently  
\* consists of 15 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1289: contig of 1289 bp in length  
\* 1290 1389: gap of unknown length  
\* 1390 3596: contig of 2207 bp in length  
\* 3597 3696: gap of unknown length  
\* 3697 5665: contig of 1969 bp in length  
\* 5666 5765: gap of unknown length  
\* 5766 9155: contig of 3389 bp in length  
\* 9156 9254: gap of unknown length  
\* 9255 14184: contig of 4930 bp in length  
\* 14185 14285: gap of unknown length  
\* 14286 19239: contig of 4955 bp in length  
\* 19240 19339: gap of unknown length  
\* 19340 25026: contig of 5687 bp in length  
\* 25027 25126: gap of unknown length

FEATURES  
source  
1. 247074  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="UNK"  
/clone="RP24-298L8"

misc\_feature  
1. 1289  
/note="assembly\_name:Contig15"

misc\_feature  
1390. 3596  
/note="assembly\_name:Contig16"

misc\_feature  
3697. 5665  
/note="assembly\_name:Contig17"

misc\_feature  
5766. 9154  
/note="assembly\_name:Contig18"

misc\_feature  
9255. 14184  
/note="assembly\_name:Contig19"

misc\_feature  
14285. 19239  
/note="assembly\_name:Contig20"

misc\_feature  
19340. 25026  
/note="assembly\_name:Contig21"

misc\_feature  
25127. 35457  
/note="assembly\_name:Contig22"

misc\_feature  
3558. 45385  
/note="assembly\_name:Contig23"

misc\_feature  
45486. 60818  
/note="assembly\_name:Contig24"

misc\_feature  
60919. 79229  
/note="assembly\_name:Contig25"

misc\_feature  
79330. 104699  
/note="assembly\_name:Contig26"

misc\_feature  
104800. 133578  
/note="assembly\_name:Contig27"

misc\_feature  
133679. 171553  
/note="assembly\_name:Contig28"

misc\_feature  
171654. 247074  
/note="assembly\_name:Contig29"

BASE COUNT 66398 a 54384 c 55492 g 63387 t 1413 others

ORIGIN  
Query Match 1.2% Score 22; DB 2; Length 247074;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 979 AGTTGTCATTAACCTACAGGAA 1000  
Db 202271 AGTTGTCATTAACCTACAGGAA 202292

RESULT 11  
AC120685/c 293272 bp DNA linear HTG 24-AUG-2002  
LOCUS Rattus norvegicus clone CH230-13J13, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION  
AC120685  
\*\*\*, 118 unordered pieces.  
AC120685  
AC120685.3 GI:22164888  
HTG; HTGS\_PHASE1.  
KEYWORDS  
Norway rat.  
SOURCE  
ORGANISM  
Rattus norvegicus



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

#### REFERENCE AUTHORS

1 (bases 1 to 293272)  
Muzny, D., Marie, E., Metzger, M., Lee, S., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alebrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biawolo, K., Blair, C., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Chen, Z., Chu, J., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Grebgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, D., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munitadasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarpunsgaon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Pramus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sleson, I., Sitter, C.D., Smales, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

2 (bases 1 to 293272)  
Unpublished  
Direct Submission  
Worley, K.C.

3 (bases 1 to 293272)  
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Aug 9, 2002 this sequence version replaced gi:21902829.

#### COMMENT

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information

Center project name: GXNP  
Center clone name: CH230-13J13  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 180898 bases at least Q40  
Consensus quality: 191186 bases at least Q30  
Consensus quality: 198854 bases at least Q20  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 118 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1  
\* 1139: contig of 1139 bp in length  
\* 1140: gap of unknown length  
\* 1239: gap of unknown length  
\* 1240: contig of 1182 bp in length  
\* 2421: gap of unknown length  
\* 2422: gap of unknown length  
\* 2521: gap of unknown length  
\* 2522: gap of unknown length  
\* 3577: contig of 1056 bp in length  
\* 3677: gap of unknown length  
\* 3678: contig of 1668 bp in length  
\* 5346: gap of unknown length  
\* 5346: gap of unknown length  
\* 6897: contig of 1452 bp in length  
\* 6898: gap of unknown length  
\* 6998: gap of unknown length  
\* 8641: contig of 1644 bp in length  
\* 8642: gap of unknown length  
\* 8742: contig of 1320 bp in length  
\* 10062: gap of unknown length  
\* 10161: gap of unknown length  
\* 11497: contig of 1336 bp in length  
\* 11498: gap of unknown length  
\* 11597: gap of unknown length  
\* 11598: gap of unknown length  
\* 12722: contig of 1124 bp in length  
\* 12722: gap of unknown length  
\* 12822: contig of 1629 bp in length  
\* 14451: gap of unknown length  
\* 14551: gap of unknown length  
\* 15839: contig of 1289 bp in length  
\* 15939: gap of unknown length  
\* 17030: contig of 1091 bp in length  
\* 17031: gap of unknown length  
\* 17131: gap of unknown length  
\* 18367: contig of 1237 bp in length  
\* 18368: gap of unknown length  
\* 18468: contig of 1300 bp in length  
\* 19768: gap of unknown length  
\* 19868: contig of 1224 bp in length  
\* 21092: gap of unknown length  
\* 21191: gap of unknown length  
\* 22562: contig of 1371 bp in length  
\* 22562: gap of unknown length  
\* 23667: contig of 1005 bp in length  
\* 23668: gap of unknown length  
\* 25100: contig of 1333 bp in length  
\* 25101: gap of unknown length  
\* 25201: contig of 1529 bp in length  
\* 26729: gap of unknown length  
\* 26829: gap of unknown length  
\* 28078: contig of 1249 bp in length  
\* 28178: gap of unknown length  
\* 29528: contig of 1350 bp in length  
\* 29529: gap of unknown length  
\* 29629: contig of 1450 bp in length  
\* 31079: gap of unknown length  
\* 31178: contig of 1137 bp in length  
\* 32315: gap of unknown length  
\* 32415: gap of unknown length  
\* 33847: contig of 1432 bp in length  
\* 33948: gap of unknown length  
\* 35629: contig of 1682 bp in length  
\* 35729: gap of unknown length  
\* 36762: contig of 1033 bp in length  
\* 36862: gap of unknown length  
\* 38277: contig of 1415 bp in length  
\* 36863

\* 38278 38377: gap of unknown length  
\* 38378 39596: contig of 1219 bp in length  
\* 39597 39697: gap of unknown length  
\* 39697 40760: contig of 1064 bp in length  
\* 40761 40860: gap of unknown length  
\* 40861 42420: contig of 1560 bp in length  
\* 42421 42520: gap of unknown length  
\* 42521 43895: contig of 1375 bp in length  
\* 43896 43995: gap of unknown length  
\* 43996 45404: contig of 1409 bp in length  
\* 45405 47096: contig of 1592 bp in length  
\* 47097 47197: gap of unknown length  
\* 47197 48518: contig of 1322 bp in length  
\* 48519 48619: gap of unknown length  
\* 48619 49788: contig of 1170 bp in length  
\* 49789 49888: gap of unknown length  
\* 49889 51110: contig of 1222 bp in length  
\* 51111 51210: gap of unknown length  
\* 51211 53062: contig of 1852 bp in length  
\* 53063 53162: gap of unknown length  
\* 53163 54912: contig of 1750 bp in length  
\* 54913 55012: gap of unknown length  
\* 55013 56564: contig of 1552 bp in length  
\* 56565 56664: gap of unknown length  
\* 56665 57895: contig of 1231 bp in length  
\* 57896 57995: gap of unknown length  
\* 57996 59279: contig of 1284 bp in length  
\* 59280 61376: gap of unknown length  
\* 61377 61476: gap of unknown length  
\* 61477 63550: contig of 2074 bp in length  
\* 63551 63650: gap of unknown length  
\* 63651 65479: contig of 1829 bp in length  
\* 65480 65579: gap of unknown length  
\* 65580 67124: contig of 1545 bp in length  
\* 67125 67224: gap of unknown length  
\* 67225 68818: contig of 1594 bp in length  
\* 68819 68918: gap of unknown length  
\* 68919 70287: contig of 1369 bp in length  
\* 70288 70387: gap of unknown length  
\* 70388 71972: contig of 1584 bp in length  
\* 71972 72071: gap of unknown length  
\* 72072 73921: contig of 1850 bp in length  
\* 73922 74021: gap of unknown length  
\* 74022 75243: contig of 1222 bp in length

Query Match 1.2%; Score 22; DB 2; Length 293272;  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 979 AGTGTGCATAACTACAGAA 1000  
Db 265159 AGTGTGCATAACTACAGAA 265138

RESULT 12  
AF026147  
LOCUS  
DEFINITION  
AF026147 15389 bp DNA linear BCT 30-MAY-1998  
AF026147  
Bacillus subtilis YojA (yojA), YojB (yojB), YojC (yojC), YojD (yojD), YojE (yojE), YojF (yojF), YojG (yojG), YojH (yojH), YojI (yojI), YojJ (yojJ), YojK (yojK), YojL (yojL), YojM (yojM), YojN (yojN), and YojO (yojO) genes, complete cds; and OdaA (odaA) gene, partial cds.

ACCESSION  
AF026147.1 GI:3169316  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Bacillus subtilis.  
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE  
1 (bases 1 to 15389)  
Park, S.-H., Shin, B.-S., Choi, S.-K., and Ghim, S.-Y.  
AUTHORS  
TITLE  
DNA sequences of a 15.4 kb fragment covering the 181 degree region

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
of the Bacillus subtilis genome  
Unpublished  
2 (bases 1 to 15389)  
Park, S.-H., Shin, B.-S., and Choi, S.-K.  
Direct Submission  
Submitted (22-SEP-1997) Applied Microbiology Research Division,  
Korea Research Institute of Bioscience and Biotechnology, P.O.Box  
115, Yusong, Taejeon 305-600, Korea  
Location/Qualifiers  
1.15389  
/organism="Bacillus subtilis"  
/strain="168"  
/db\_xref="taxon:1423"  
/map="181 degrees"  
/note="sequence upstream of the odaA gene"  
7.1341  
/gene="yojA"  
7.1341  
/note="similar to Bacillus subtilis gluconate permease  
encoded by the sequence presented in GenBank Accession  
Number J02584"  
/codon\_start=1  
/transl\_table=1  
/product="YojA"  
/protein\_id="AAC17849.1"  
/db\_xref="GI:3169317"  
/translation="MPILIVANGVILFLIKVKLNTFVSLIVSVFVAIGLDIN  
KVLISLETIGGQLGHLAIVFGIAMLGLVSDAGGYRIATLLDKRKRIOAAV  
VASFLIGLIFPEVGLVLIPIVIAIAKELKMPFLYLGIPMAAOLNTHGFLPHAP  
TAISVAYGHAIGOVLFGLIIIAVPTTVIAGLPFNKAMKRFPAVOKGNLSGLPRK  
EROLDETPGFAISATSLPVIEMAMATFSLISEHSGKDIIEFIGPTAMLSL  
LLAATMGVARKISMOETRCISSEISOTAMLLIIGGGARFQVILDGVDYADF  
SVIACHVNDAGVMTVAIVRLTSGATVAALTTAAMAPIMAGSVNPLVNLGAG  
1426.1662  
/gene="yojB"  
1426.1662  
/gene="yojB"  
/function="unknown"  
/codon\_start=1  
/transl\_table=1  
/product="YojB"  
/protein\_id="AAC17850.1"  
/db\_xref="GI:3169318"  
/translation="MPPHSYRGLIPGPGYPARBPFLGALVGGLLGFLGSLFN  
YSRPVAVPYGGYGGPYGFAGVPGYGPY"  
1770.2171  
/gene="yojC"  
complement(1770..2171)  
/function="unknown"  
/codon\_start=1  
/transl\_table=1  
/product="YojC"  
/protein\_id="AAC17851.1"  
/db\_xref="GI:3169319"  
/translation="MPENKRYPPHNOSEKALKRADPKREVEYVNRUNESVGSNYA  
ADPFRDPLDQKHPAKDPKDPKVDIDIFETADHVFVKVPISEKMLQVAKATSH  
EIMLEMLPRAESEKSPSLFSQOTNNKSRLL"  
2270.3061  
/gene="yojD"  
2270.3061  
/gene="yojD"  
/function="unknown"  
/note="similar to Escherichia coli RaxD protein;  
Swiss-Prot Accession Number P27844"  
/codon\_start=1  
/transl\_table=1  
/product="YojD"  
/protein\_id="AAC17852.1"  
/db\_xref="GI:3169320"  
/translation="MQHAETKGVITYTAVSFIMWGLFPLYWKLLEQLPALDLIAHRI  
MSFPMCIIVLFLRQKIGIQMQLSRKNGGILSLFLASILISINFFVIYIAVNHGFL



## COMMENT

Submitted by:

Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63110, USA  
e-mail: jsplith@wustl.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

## FEATURES

Location/Qualifiers  
1..41830

/organism="Caenorhabditis briggsae"  
/strain="Gujarat G16"  
/db\_xref="taxon:6238"  
/clone="G4E08"

BASE COUNT 12651 a 8319 c 7746 g 13114 t

## ORIGIN

Query Match 1.1%; Score 21; DB 3; Length 41830;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1651 TTTTGCAATTATATGACCAT 1671  
Db 6982 TTTTGCAATTATATGACCAT 6962

## RESULT 14

AL355552/c 90157 bp DNA linear PRI 16-FEB-2001  
LOCUS Human DNA sequence from clone RP11-217F17 on chromosome 13 contains  
DEFINITION a ribosomal protein L7 pseudogene, 5'UTRs and 3'UTRs, complete  
sequence.

ACCESSION AL355552  
VERSION AL355552.8 GI:11121041

## KEYWORDS

HTG.

## SOURCE

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 90:57)

Smith, M.

Direct Submission

Submitted (09-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requesters: clonerequest@sanger.ac.uk

On Nov 8, 2000 this sequence version replaced gi:11062390.

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 13, constructed by the Sanger Centre Chromosome 13  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr13

IMPORTANT: This sequence is not the entire insert of clone  
RP11-217F17. It may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.  
The true right end of clone RP11-217F17 is at 90157 in this  
sequence. The true left end of clone RP11-562B17 is at 56085 in  
this sequence. The true right end of clone RP11-155N20 is at 100 in  
this sequence. This sequence was finished as follows unless  
otherwise noted: all regions were either double-stranded or  
sequenced with an alternate chemistry or covered by high quality

## FEATURES

source

data (i.e., phred quality >= 30); an attempt was made to resolve  
all sequencing problems, such as compressions and repeats; all  
regions were covered by at least one plasmid subclone or more than  
one M13 subclone; and the assembly was confirmed by restriction  
digest. RP11-217F17 is from the library RPCI-11.1 constructed by  
the group of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6.

Location/Qualifiers  
1..90157

/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/chromosome="13"

/clone="RP11-217F17"

/clone\_lib="RPCI-11.1"

complement(1..86)

/note="match: GSS: Em:AQ386890"

complement(1147..1651)

/note="match: GSS: Em:AQ522917"

2147..2490

/note="THE1B repeat: matches 1..364 of consensus"

2571..2754

/note="MER5A repeat: matches 2..186 of consensus"

2827..2892

/note="MER5A repeat: matches 118..186 of consensus"

4341..4631

/note="L2 repeat: matches 690..988 of consensus"

4630..5169

/note="L2 repeat: matches 2165..2750 of consensus"

5195..5510

/note="AL15X repeat: matches 1..312 of consensus"

6835..7324

/note="match: GSS: Em:AQ433386"

7814..8054

/note="L1M3 repeat: matches 5792..6037 of consensus"

8057..8256

/note="L1M3 repeat: matches 5..194 of consensus"

8257..8284

/note="L1M3 repeat: matches 5..194 of consensus"

8511..8644

/note="L1M3 repeat: matches 617..741 of consensus"

8645..9755

/note="L1M3 repeat: matches 273..1383 of consensus"

9756..10048

/note="L1M3 repeat: matches 273..1383 of consensus"

10049..10590

/note="L1M3 repeat: matches 1..293 of consensus"

10591..10903

/note="L1M3 repeat: matches 1383..1927 of consensus"

10904..11055

/note="L1M3 repeat: matches 1..311 of consensus"

12018..12331

/note="L1M3 repeat: matches 1927..2078 of consensus"

15295..15893

/note="L1M3 repeat: matches 1..312 of consensus"

16122..16719

/note="L1M2 repeat: matches 19..572 of consensus"

16720..16958

/note="MER41B repeat: matches 1..606 of consensus"

16959..17373

/note="L1P1A6 repeat: matches 5944..6165 of consensus"

17093..17373

/note="L2 repeat: matches 1539..1818 of consensus"

17376..17653

/note="L1M3 repeat: matches 117..389 of consensus"

17872..18195

/note="L1M3 repeat: matches 1..288 of consensus"

18202..18501

/note="L1M3 repeat: matches 1..303 of consensus"

19639..19678

/note="L1M3 repeat: matches 1..303 of consensus"

19645..19676

/note="8 copies 4 mer tctg 100% conserved"

19680..19731

/note="13 copies 4 mer atat 73% conserved"

repeat\_region 19758..19976  
/note="AluDb repeat: matches 102. .287 of consensus"  
repeat\_region 20823..20866  
/note="11 copies 4 mer tcct 79% conserved"  
repeat\_region 20983..21110  
/note="64 copies 2 mer ta 90% conserved"  
repeat\_region 21326..21616  
/note="MER46B repeat: matches 1. .236 of consensus"  
repeat\_region 21753..21930  
/note="MIR repeat: matches 31. .207 of consensus"  
repeat\_region 22354..22551  
/note="MER2 repeat: matches 3. .210 of consensus"  
repeat\_region 23293..23566  
/note="AluY repeat: matches 39. .311 of consensus"  
repeat\_region 25962..26019  
/note="2 copies 29 mer 96% conserved"  
repeat\_region 27578..27617  
/note="10 copies 4 mer agag 90% conserved"  
repeat\_region 29695..29930  
/note="MIR repeat: matches 20. .245 of consensus"  
repeat\_region 31437..31583  
/note="MLT1-INTERNAL repeat: matches 6. .150 of consensus"  
repeat\_region 31594..32041  
/note="L1P29 repeat: matches 5703. .6159 of consensus"  
repeat\_region 32042..32416  
/note="MLT1-INTERNAL repeat: matches 150. .639 of consensus"  
repeat\_region 33079..33102  
/note="12 copies 2 mer tt 100% conserved"  
repeat\_region 33488..33899  
/note="MLT1B repeat: matches 1. .384 of consensus"  
repeat\_region 34181..34229  
/note="MIR repeat: matches 21. .71 of consensus"  
repeat\_region 35227..35406  
/note="2 copies 90 mer 85% conserved"  
repeat\_region 36236..36387  
/note="MIR repeat: matches 20. .174 of consensus"  
misc\_feature Complement(37604..38241)  
misc\_feature /note="match: GSS: Em:AQ357261"  
repeat\_region 37654..38484  
/note="match: GSS: Em:AQ750273"  
repeat\_region 39395..39472  
/note="39 copies 2 mer tt 64% conserved"  
repeat\_region 40268..40583  
/note="AluX repeat: matches 1. .312 of consensus"  
repeat\_region 40687..40850  
/note="MERS8D repeat: matches 1. .212 of consensus"  
repeat\_region 40881..40914  
/note="MERS8D repeat: matches 164. .197 of consensus"  
repeat\_region 41288..41553  
/note="MLT1E repeat: matches 307. .568 of consensus"  
repeat\_region 41582..41751  
/note="MERS8D repeat: matches 269. .431 of consensus"  
misc\_feature complement(42592..43084)  
/note="match: STS: Em:G56777"  
repeat\_region 44442..44948  
/note="L2 repeat: matches 2151. .2690 of consensus"  
repeat\_region 45948..46081  
/note="MIR repeat: matches 38. .172 of consensus"  
repeat\_region 46286..46468  
/note="MIR repeat: matches 48. .262 of consensus"  
misc\_feature complement(47295..47695)  
/note="match: GSS: Em:AQ347277"  
misc\_feature match: STS: Em:G57525"  
misc\_feature 47698..48271  
/note="match: GSS: Em:AQ779313"  
misc\_feature 48610..48816  
/note="match: GSS: Em:B54651"  
misc\_feature complement(48612..48788)  
/note="match: GSS: Em:B51274"  
gene 48647..49389  
/gene="BA217F17.1"

CDS 48647..49389  
/gene="BA217F17.1"  
/note="bA217F17.1 (ribosomal protein L7 pseudogene)"  
match: proteins: Sw:Q12213 Sw:P25457 Tr:Q92U42 Sw:P11874  
Sw:P17937 Sw:Q42208 Sw:P05737 Tr:Q92NU7"  
/pseudo  
/codon\_start=1  
/evidence=not\_experimental  
misc\_feature complement(48659..48765)  
/note="match: GSS: Em:AQ057230"  
repeat\_region 49847..50186  
/note="MER6 repeat: matches 54. .399 of consensus"  
Query Match 1.1%; Score 21; DB 9; Length 90157;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 537 TTAAGAAATCTGTAGAAAAG 557  
|||||  
Db 43199 TTAAGAAATCTGTAGAAAAG 43179  
|||||  
RESULT 15  
AC000066 AC000066 94960 bp DNA linear PRI 21-DEC-1999  
LOCUS Homo sapiens BAC clone CTA-293F11 from 7q21-7q22, complete  
DEFINITION sequence.  
AC000066  
AC000066 AC000066.1 GI:3645943  
VERSION  
KEYWORDS HYG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 94960)  
AUTHORS Sulston, J.E. and Waterston, R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 99063792  
PUBMED 9847074  
REFERENCE 2 (bases 1 to 94960)  
AUTHORS Wu, X., Graves, T. and Bradshaw, H.  
TITLE The sequence of Homo sapiens BAC clone CTA-293F11  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 94960)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (13-NOV-1996) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 4 (bases 1 to 94960)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (24-SEP-1998) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 5 (bases 1 to 94960)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (17-FEB-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 6 (bases 1 to 94960)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT On Sep 24, 1998 this sequence version replaced gi:1669365.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
----- Center project name: H\_RG293F11  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7> or <mailto:egreen@nhgri.nih.gov>

SOURCE INFORMATION:  
Clone CTA-293F11 is from a release of the human BAC library CITR-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).  
VECTOR: pBelOBAc11  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is CTR-104F4, 200 bp overlap; the clone sequenced to the right is GSI-541B18, 200 bp overlap. Actual start of this clone is unknown; actual end is at 94960 of RG293F11.

Location/Qualifiers

1..94960

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="7"

/map="7q21-7q22"

/clone="CTA-293F11"

/clone\_lib="CITR-HS-A"

repeat\_region

2..96

/rpt\_family="MER1\_type"

repeat\_region

883..990

/rpt\_family="L1"

repeat\_region

991..1234

/rpt\_family="Alu"

repeat\_region

1421..1722

/rpt\_family="Alu"

misc\_feature

2303..2340

/note="match to EST AA329389 (NID:g1981781)"

misc\_feature

2343..2630

/note="match to EST AA329389 (NID:g1981781)"

repeat\_region

4161..4176

/rpt\_family="Alu"

repeat\_region

4177..4203

/rpt\_family="Alu"

repeat\_region

4204..4480

/rpt\_family="(TAA)n"

repeat\_region

4605..4871

/rpt\_family="Alu"

repeat\_region

4892..4935

/rpt\_family="L1"

repeat\_region

5906..6100

/rpt\_family="L2"

repeat\_region

6316..6367

/rpt\_family="L2"

repeat\_region

6368..6391

/rpt\_family="MIR"

repeat\_region

6469..6779

/rpt\_family="L2"

repeat\_region

6865..7757

/rpt\_family="Alu"

repeat\_region

STS  
complement(8037..8323)  
/db\_xref="GI:940344"

repeat\_region

8193..8215

/rpt\_family="AT\_rich"

repeat\_region

8612..8641

/rpt\_family="AT\_rich"

repeat\_region

8785..8830

/rpt\_family="(CA)n"

repeat\_region

8833..9119

/rpt\_family="Alu"

repeat\_region

9792..9907

/rpt\_family="L1"

repeat\_region

10207..11339

/rpt\_family="L1"

repeat\_region

11362..11489

/rpt\_family="Alu"

repeat\_region

11494..11762

/rpt\_family="L1"

repeat\_region

11764..11907

/rpt\_family="L1"

repeat\_region

11908..13266

/rpt\_family="L1"

repeat\_region

13267..13545

/rpt\_family="Alu"

repeat\_region

13546..14169

/rpt\_family="L1"

repeat\_region

14170..14490

/rpt\_family="Alu"

repeat\_region

14491..16242

/rpt\_family="L1"

repeat\_region

16395..16858

/rpt\_family="L1"

repeat\_region

16859..17170

/rpt\_family="Alu"

repeat\_region

17171..17233

/rpt\_family="L1"

repeat\_region

17691..17989

/rpt\_family="L1"

repeat\_region

17996..18072

/rpt\_family="Mariner"

repeat\_region

18181..18249

/rpt\_family="L1"

repeat\_region

18401..18692

/rpt\_family="Alu"

gene

<19022..>86200

/gene="WUSC:H\_RG293F11.1"

join(<19022..19270,25591..25635,37460..37513,38187..38357,

39923..40078,40905..41102,46150..48537,57731..57944,

59551..59630,61431..61569,62319..62404,67540..67654,

68116..68311,75197..75293,76814..76906,83721..84074,

85976..>86200)

/gene="WUSC:H\_RG293F11.1"

/note="match to AF026245 (PID:92623068); but note

differences at carboxy terminus; H\_RG293F11.1"

/codon\_start=1

/product="Yotiao"

/protein\_id="AAC60380.1"

/db\_xref="GI:3645944"

/translation="PRKRAQSDGSPKQKKRKTSSKHVSAHDLNDQSCN

EMYNISQVSTVPESTIMRTLSGITSBQGSFVLSSTTDDCSSENGC

SPVMTGKPTNLAREEGRVDYSSEQAQSPHLEMBEELACKQHEIEINRELE

EMRVYGTGEGLOQLQEFPAIKQDGIITQLPANQQRKDETFWRFLELTESQK

LQIQOQASSETLANSSTADLLQAKQIILHQQLLEQDHLLEDTYQKKEDFT

MQISFLQKRIYVEMEDQKVENSKKEIIOKETIIEBLNFKIIEEKTELKDKLT

TADKLGLQLOIYQKQEIKNMKLEITNSKQSESEIKQLMGVTEBLQKRNKD

SOETDIYVMEQETQRLQRLAELDEMGQIYQKQELIRQMAEMEMKTRPKG

EMENLRYSNITNEDQIKMNVAINELNFKLQDTNQCKEKLELQILIEKKALQ

ROLEDLVELSPREIQRQRTIABQSKINLNAKSLSTVEDLKAETVASESKEL

ELKHEAVTNFKIKLEMLEKERNVLDMASSQEALETQLFTEHEELSKLED

LEIHRINIEKLNKIGIHYQOQIDGLQNMSSQKIEITQFKNDLITQNDLIEISK

LKDLQQSLVNSKSEMTQIINLEQKEIILROEKEKGTLDQVOELQKTELEKOM

KEKENDQEKPAQLAENSILKDEKTELDLKHITPQSEERLLFDLSISKSDSV

WEKEIEILLIENEDLKQOCIQLNNEIEKQRTSPFAKNEFVNVOELQBEYACLLKVK  
 DDLEDSKQKELEYKSLKALNELHLQRLNPTTVKMKSSVDFEDDKTFVATLEMGV  
 VKDXTDLEIKLEVKLEQLSRLDLSEQLKQKHGEISFLNEEVKSLKQKEQVS  
 LRCRELIINHNAENVQCDTQVSSLLDGVVTMTSRGAEGSVKVKSGFGEESKIM  
 VEDKVPENVTGVEESKQEQCLILDHLPVTKESSLRATQPSNDKQLKQLNLVKSEQN

Query Match 1.1%; Score 21; DB 9; Length 94960;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1395 ATTCTCTTTTGAAGAAATTC 1415  
 |||||  
 Db 81286 ATTCTCTTTTGAAGAAATTC 81306

RESULT 16  
 AL161743/c  
 LOCUS AL161743 100170 bp DNA linear PRI 09-APR-2001  
 DEFINITION Human DNA sequence from clone RP4-730B13 on chromosome 1q42.11-42.13  
 Contains part of the DISC1 (disrupted in schizophrenia 1) gene,  
 STSs and GSSs, complete sequence.

ACCESSION AL161743  
 VERSION AL161743.20 GI:11121016  
 KEYWORDS HTG; DISC1; schizophrenia.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 100170)  
 Coville, G.

REFERENCE Direct Submission  
 Submitted (16-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk

On Nov 8, 2000 this sequence version replaced gi:10715861.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information  
 on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
 Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr1  
 This sequence is the entire insert of clone RP4-730B13 This  
 sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. RP4-730B13 is from  
 the library RPCI-4 constructed by the group of Pieter de Jong. For  
 further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pCYPAC2.

## FEATURES

## source

Location/Qualifiers  
 1..100170  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="q42.11-42.13"  
 /clone="RP4-730B13"  
 /clone\_lib="RPCI-4"  
 complement(join(137..255,17641..17905,67802..67862))  
 /gene="DISC1"  
 complement(join(<137..255,17641..17905,67802..67862))

## gene

## CDS

/gene="DISC1"  
 /notes="match: cDNAs: Em:AF222980 Em:AB007926  
 match: proteins: Tr:O75045"  
 /codon\_start=3  
 /evidence=not experimental  
 /product="dj730B13.1 (disrupted in schizophrenia 1)"  
 /protein\_id="CAC36349.1"  
 /db\_xref="GI:13591447"  
 /translation="TSVKENTWYMETLKNKLCCKPLLGKWEADLEACRLLIQSL  
 QLOEARGSLSVDEDERODDLEGAAPPIPPRLHSEDKRKTPLKVLSEWKTHLIPSLHCA  
 GEOKEESYLSAELGCEKCEDIGKLLYLEDLQDLTAHSHDEDLI"  
 826..1092  
 /note="AluY repeat: matches 32..298 of consensus"  
 complement(959..1362)  
 /gene="DISC1"  
 /notes="match: GSS: Em:AQ738086"  
 complement(981..1392)  
 /gene="DISC1"  
 /notes="match: GSS: Em:B88351"  
 1093..1191  
 /notes="L2 repeat: matches 2606..2705 of consensus"  
 complement(1109..1392)  
 /gene="DISC1"  
 /note="match: GSS: Em:B71142"  
 1801..2092  
 /notes="AluJo repeat: matches 1..295 of consensus"  
 2603..2680  
 /notes="39 copies 2 mer ta 79% conserved"  
 3674..4066  
 /notes="match: STS: Em:G60191"  
 3693..3781  
 /notes="match: STS: Em:G43347"  
 4085..4381  
 /note="AluSc repeat: matches 1..298 of consensus"  
 4534..4992  
 /notes="MIR1G repeat: matches 39..512 of consensus"  
 5316..5355  
 /notes="20 copies 2 mer tg 100% conserved"  
 6320..6630  
 /notes="AluSp repeat: matches 1..313 of consensus"  
 7439..7654  
 /note="L2 repeat: matches 2428..2710 of consensus"  
 12707..13261  
 /notes="MER39 repeat: matches 3..549 of consensus"  
 complement(13490..14103)  
 /gene="DISC1"  
 /note="match: GSS: Em:AQ041978"  
 14437..14548  
 /notes="LTR34 repeat: matches 1..116 of consensus"  
 complement(14853..15183)  
 /gene="DISC1"  
 /notes="match: GSS: Em:AQ542245"  
 14925..15148  
 /notes="LIME1 repeat: matches 5931..6150 of consensus"  
 15195..15594  
 /notes="match: GSS: Em:AQ168058"  
 15638..16158  
 /notes="MER39 repeat: matches 6..543 of consensus"  
 17283..17372  
 /notes="10 copies 9 mer tataataa 68% conserved"  
 18189..18382  
 /notes="L2 repeat: matches 2298..2496 of consensus"  
 18477..18525  
 /note="L2 repeat: matches 2601..2649 of consensus"  
 18533..18908  
 /notes="THE1C repeat: matches 1..371 of consensus"  
 18909..20484  
 /notes="THE1C-internal repeat: matches 35..1580 of  
 consensus"  
 20804..21002  
 /notes="LIMEC repeat: matches 1217..1427 of consensus"  
 21033..21399  
 /notes="LIMEC repeat: matches 1497..1881 of consensus"

## repeat\_region

## misc\_feature

## misc\_feature

## repeat\_region

## misc\_feature

## repeat\_region

## repeat\_region

## misc\_feature

## misc\_feature

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## misc\_feature

## repeat\_region

## misc\_feature

## repeat\_region

## misc\_feature

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

```

repeat_region 21549..21786
/note="119 copies 2 mer ag 56% conserved"
repeat_region 21828..21951
/note="L2 repeat: matches 966..1088 of consensus"
repeat_region 21952..22071
/note="L1MC/D repeat: matches 5371..5496 of consensus"
repeat_region 22107..22677
/note="L1M4 repeat: matches 2922..3512 of consensus"
repeat_region 22825..23558
/note="L1M4 repeat: matches 2131..2925 of consensus"
repeat_region 23635..25145
/note="L1MD2 repeat: matches 4794..6307 of consensus"
repeat_region 25313..25611
/note="ALUSX repeat: matches 2..302 of consensus"
repeat_region 25800..26697
/note="L1MD repeat: matches 979..1932 of consensus"
repeat_region 26792..26881
/note="L1C copies 9 mer tgccttctgc 71% conserved"
repeat_region 26799..26878
/note="40 copies 2 mer tc 63% conserved"
repeat_region 27364..27547
/note="L1MD repeat: matches -17..167 of consensus"
misc_feature complement(27754..28002)
/gene="DISC1"
repeat_region 28092..28203
/note="match: GSS: Em:AQ019129"
repeat_region 30215..31339
/note="HML1 repeat: matches 626..730 of consensus"
repeat_region 31342..31377
/note="L1PA2 repeat: matches 5020..6144 of consensus"
repeat_region 31979..32041
/note="7 copies 9 mer tcatcatca 73% conserved"
repeat_region 32052..32121
/note="MTR repeat: matches 158..227 of consensus"
repeat_region 32137..32407
/note="MTR repeat: matches 1..261 of consensus"
repeat_region 32979..33095
/note="L2 repeat: matches 2593..2705 of consensus"
repeat_region 34127..34397
/note="L1ME repeat: matches 5108..5386 of consensus"
repeat_region 34495..34664
/note="L1ME repeat: matches 5567..5747 of consensus"
repeat_region 36753..36960
/note="MTR repeat: matches 35..262 of consensus"
repeat_region 37937..38298
/note="MTR1A1 repeat: matches 1..365 of consensus"
repeat_region 38332..38471
/note="MER5A repeat: matches 9..173 of consensus"
repeat_region 38492..38783
/note="ALUSX repeat: matches 1..295 of consensus"
misc_feature 39472..39939
/note="match: GSS: Em:AQ068363"
misc_feature 39501..39917
/note="match: GSS: Em:AQ088543"
misc_feature complement(39885..40383)
/gene="DISC1"
repeat_region 40529..40699
/note="MTR repeat: matches 59..233 of consensus"
repeat_region 41234..41325
/note="MTR repeat: matches 166..252 of consensus"
repeat_region 41326..41651
/note="ALUSG repeat: matches 1..311 of consensus"
repeat_region 41652..41792
/note="MTR repeat: matches 7..166 of consensus"
repeat_region 41864..42129
/note="L2 repeat: matches 2408..2688 of consensus"
repeat_region 42614..42791
/note="MTR repeat: matches 32..211 of consensus"
repeat_region 42877..43020

```

1.1%; Score 21; DB 9; Length 100170;

```

Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1400 TTTTGAAGAAATTCATA 1420
Db 21386 TTTTGAAGAAATTCATA 21366

```

```

RESULT 17
AC122235_2/c
WPCOMMENT
Sequence split into 5 fragments
Fragment Name Begin End
AC122235_1 1 110000
AC122235_2 100001 210000
AC122235_3 200001 310000
AC122235_4 300001 410000
AC122235_5 400001 434777
Continuation (3 of 5) of AC122235 from base 200001 (AC122235 Mus musculus chromosome UR

```

```

Query Match 1.1%; Score 21; DB 2; Length 110000;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 992 CTACAGAGCTGCTCTTCT 1012
Db 83906 CTACAGAGCTGCTCTTCT 83886

```

```

RESULT 18
AC124954/c
LOCUS
DEFINITION
Medicago truncatula clone mth1-10g9, WORKING DRAFT SEQUENCE, 5
ordered pieces.
AC124954
AC124954.5 GI:21717195
VERSION
HTG: HTGS PHASE2; HTGS_DRAFT.
KEYWORDS
barrel medic.
SOURCE
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 114741)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D., and Roe,B.A.
Medicago truncatula BAC Clone mth1-10g9
Unpublished
2 (bases 1 to 114741)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D., and Roe,B.A.
Direct Submission
Submitted (20-JUN-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 114741)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D., and Roe,B.A.
Direct Submission
Submitted (15-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Jul 9, 2002 this sequence version replaced gi:21672150.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have



1. (bases 1 to 123091)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaratingue, H.C., Are, J.R., Ayele, M., Banks, T., Barbara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieval, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A. L., Ding, Y., Din, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsí, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kurehi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegé, H., Lozadó, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogwu, M., Okwuonu, G.,

```

* 18592 18691: gap of unknown length
* 18692 20268: contig of 1577 bp in length
* 20269 20368: gap of unknown length
* 20369 21874: contig of 1506 bp in length
* 21875 21974: gap of unknown length
* 21975 23378: contig of 1404 bp in length
* 23379 23478: gap of unknown length
* 23479 24606: contig of 1128 bp in length
* 24607 24706: gap of unknown length
* 24707 26033: contig of 1327 bp in length
* 26034 27532: gap of unknown length
* 27533 27632: contig of 1399 bp in length
* 27633 29257: gap of unknown length
* 29258 29357: gap of unknown length
* 29358 30896: contig of 1539 bp in length
* 30897 30996: gap of unknown length
* 30997 32545: contig of 1549 bp in length
* 32546 32645: gap of unknown length
* 32646 33774: contig of 1123 bp in length
* 33775 33874: gap of unknown length
* 33875 35039: contig of 1215 bp in length
* 35040 35139: gap of unknown length
* 35140 37444: contig of 2255 bp in length
* 37445 37544: gap of unknown length
* 37545 39032: contig of 1538 bp in length
* 39033 39132: gap of unknown length
* 39133 40532: contig of 1350 bp in length
* 40533 40632: gap of unknown length
* 40633 41737: contig of 1075 bp in length
* 41708 41807: gap of unknown length
* 41808 43935: contig of 2188 bp in length
* 43936 44035: gap of unknown length
* 44036 45348: contig of 1253 bp in length
* 45349 45448: gap of unknown length
* 45449 47732: contig of 2344 bp in length
* 47733 47832: gap of unknown length
* 47833 49418: contig of 1526 bp in length
* 49419 49518: gap of unknown length
* 49519 51246: contig of 1728 bp in length
* 51247 51346: gap of unknown length
* 51347 52750: contig of 1404 bp in length
* 52751 52850: gap of unknown length
* 52851 54709: contig of 1859 bp in length
* 54710 54809: gap of unknown length
* 54810 56706: contig of 1897 bp in length
* 56707 56806: gap of unknown length
* 56807 59494: contig of 2688 bp in length
* 59495 59594: gap of unknown length
* 59595 60822: contig of 1228 bp in length
* 60823 60922: gap of unknown length
* 60923 62056: contig of 1134 bp in length
* 62057 62156: gap of unknown length
* 62157 64938: contig of 2782 bp in length
* 64939 65038: gap of unknown length
* 65039 66239: contig of 1261 bp in length
* 66240 66300: gap of unknown length
* 66301 68146: contig of 1747 bp in length
* 68147 68246: gap of unknown length
* 68247 69841: contig of 1595 bp in length
* 69842 69941: gap of unknown length
* 69942 71777: contig of 1836 bp in length
* 71778 71877: gap of unknown length
* 71879 73269: contig of 1392 bp in length
* 73270 73369: gap of unknown length
* 73370 75564: contig of 2195 bp in length
* 75565 75664: gap of unknown length
* 75665 77306: contig of 1642 bp in length
* 77307 77406: gap of unknown length
* 77407 79833: contig of 2427 bp in length
* 79834 82424: gap of unknown length
* 82425 82524: gap of unknown length

```

```

Query Match      1.1%; Score 21; DB 2; Length 123091;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1403 TTGAAGAAATTCATAGAT 1423
Db 47571 TTGAAGAAATTCATAGAT 47591

RESULT 20
AC127017/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-66a17, *** SEQUENCING IN PROGRESS
ACCESSION
AC127017
VERSION
AC127017.1 GI:21734917
KEYWORDS
HTG; HTGS PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 124028)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaraltinge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbarta,J., Benson,J., Bimagne,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinn,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Dublin,K.J.,
Fairhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flegg,N., Ford,U., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Hornai,F., Howard,S., Huber,J., Hulix,S., Hune,U., Jackson,L.B.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,Y., Louisedge,H.,
Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Ruiz,M., Ren,Y.,
Rives,M., Rojas,A., RojudoKan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sison,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,

```

Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G. and Gibbs,R.  
 Direct Submision  
 2 (bases 1 to 124028)  
 Worley,K.C.  
 Direct Submision  
 Submitted (12-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 124028)  
 Worley,K.C.  
 Direct Submision  
 Submitted (31-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 ----- Genome Center of Medicine  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: RAWX  
 Center clone name: CH230-66A17  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100k of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 66056 bases at least Q40  
 Consensus quality: 71270 bases at least Q30  
 Consensus quality: 74172 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 69 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 1003: contig of 1003 bp in length  
 \* 1004 1103: gap of unknown length  
 \* 1104 2256: contig of 1153 bp in length  
 \* 2257 2356: gap of unknown length  
 \* 2357 3434: contig of 1078 bp in length  
 \* 3435 3534: gap of unknown length  
 \* 3535 4606: contig of 1072 bp in length  
 \* 4607 4706: gap of unknown length  
 \* 4707 5786: contig of 1080 bp in length  
 \* 5787 5886: gap of unknown length  
 \* 5887 7370: contig of 1484 bp in length  
 \* 7371 7470: gap of unknown length  
 \* 7471 8841: contig of 1371 bp in length  
 \* 8842 8941: gap of unknown length  
 \* 8942 10117: contig of 1176 bp in length  
 \* 10118 10217: gap of unknown length  
 \* 10218 11272: contig of 1055 bp in length  
 \* 11273 11372: gap of unknown length  
 \* 11373 12418: contig of 1046 bp in length  
 \* 12419 12518: gap of unknown length  
 \* 12519 14138: contig of 1620 bp in length  
 \* 14139 14238: gap of unknown length  
 \* 14239 15243: contig of 1005 bp in length  
 \* 15244 15343: gap of unknown length  
 \* 15344 16432: contig of 1089 bp in length  
 \* 16433 16532: gap of unknown length  
 \* 16533 17608: contig of 1076 bp in length  
 \* 17609 17708: gap of unknown length  
 \*  
 \* 17709 18850: contig of 1141 bp in length  
 \* 18850 20015: gap of unknown length  
 \* 20015 20116: gap of unknown length  
 \* 20116 21605: contig of 1490 bp in length  
 \* 21605 21705: gap of unknown length  
 \* 21705 23231: contig of 1526 bp in length  
 \* 23231 24853: contig of 1522 bp in length  
 \* 24853 26081: gap of unknown length  
 \* 26081 26181: contig of 1128 bp in length  
 \* 26181 27398: contig of 1217 bp in length  
 \* 27398 27498: gap of unknown length  
 \* 27498 28574: contig of 1076 bp in length  
 \* 28574 28674: gap of unknown length  
 \* 28674 29750: contig of 1076 bp in length  
 \* 29750 29850: gap of unknown length  
 \* 29850 31004: contig of 1154 bp in length  
 \* 31004 32493: gap of unknown length  
 \* 32493 32593: contig of 1389 bp in length  
 \* 32593 33769: gap of unknown length  
 \* 33769 33869: contig of 1176 bp in length  
 \* 33869 35109: gap of unknown length  
 \* 35109 35209: contig of 1240 bp in length  
 \* 35209 36997: gap of unknown length  
 \* 36997 37097: contig of 1788 bp in length  
 \* 37097 38206: gap of unknown length  
 \* 38206 38306: contig of 1109 bp in length  
 \* 38306 39463: gap of unknown length  
 \* 39463 39563: contig of 1157 bp in length  
 \* 39563 41966: gap of unknown length  
 \* 41966 42066: contig of 2403 bp in length  
 \* 42066 43357: gap of unknown length  
 \* 43357 43457: contig of 1291 bp in length  
 \* 43457 45525: gap of unknown length  
 \* 45525 45625: contig of 2068 bp in length  
 \* 45625 47806: gap of unknown length  
 \* 47806 47906: contig of 2181 bp in length  
 \* 47906 49278: gap of unknown length  
 \* 49278 49379: contig of 1373 bp in length  
 \* 49379 51108: gap of unknown length  
 \* 51108 51209: contig of 1729 bp in length  
 \* 51209 52355: gap of unknown length  
 \* 52355 52455: contig of 1147 bp in length  
 \* 52455 54114: gap of unknown length  
 \* 54114 54214: contig of 1659 bp in length  
 \* 54214 55878: gap of unknown length  
 \* 55878 55978: contig of 1664 bp in length  
 \* 55978 57262: gap of unknown length  
 \* 57262 57363: contig of 1284 bp in length  
 \* 57363 59733: gap of unknown length  
 \* 59733 59833: contig of 2371 bp in length  
 \* 59833 62133: gap of unknown length  
 \* 62133 62233: contig of 2300 bp in length  
 \* 62233 63745: gap of unknown length  
 \* 63745 63845: contig of 1512 bp in length  
 \* 63845 65057: gap of unknown length  
 \* 65057 65157: contig of 1212 bp in length  
 \* 65157 66293: gap of unknown length  
 \* 66293 66394: contig of 1136 bp in length  
 \* 66394 68088: gap of unknown length  
 \* 68088 68188: contig of 1695 bp in length  
 \* 68188 69189: gap of unknown length  
 \* 69189 69289: contig of 1001 bp in length  
 \* 69289 70839: gap of unknown length  
 \* 70839 72523: contig of 1550 bp in length  
 \* 72523 72623: gap of unknown length  
 \* 72623 74389: contig of 1584 bp in length  
 \* 74389 74489: gap of unknown length  
 \* 74489 75927: contig of 1766 bp in length  
 \* 75927 75992: gap of unknown length  
 \* 75992 75997: contig of 1438 bp in length

```

* 75928 76027: gap of unknown length
* 76028 77665: contig of 1538 bp in length
* 77666 77665: gap of unknown length
* 77666 79731: contig of 2066 bp in length
* 79732 79831: gap of unknown length
* 79832 83121: contig of 3290 bp in length
* 83122 83221: gap of unknown length

Query Match 1.1%; Score 21; DB 2; Length 124028;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1221 TAAGCAGCAACTATCGAAGA 1241
|||||
DB 45185 TAAGCAGCAACTATCGAAGA 45165

RESULT 21
LOCUS AL445200 133054 bp DNA linear HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome 1 clone RP5-852F22, *** SEQUENCING IN
ACCESSION AL445200
VERSION AL445200.1 GI:10716452
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 133054)
McLay,K.
Direct Submission
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dj852F22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 124206 bases at least Q40
Consensus quality: 129947 bases at least Q20
Consensus quality: 131534; sum-of-contigs
Insert size: 149216; 5.4% error; agarose-fp
Insert size: 149216; 5.4% error; agarose-fp
Quality coverage: 3.44x in Q20 bases; sum-of-contigs Quality
coverage: 3.12x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 6858: contig of 6858 bp in length
* 6859 6958: gap of 100 bp
* 6959 11344: contig of 4386 bp in length
* 11345 11444: gap of 100 bp
* 11445 15274: contig of 3830 bp in length
* 15275 15374: gap of 100 bp
* 15375 18330: contig of 2956 bp in length
* 18331 18430: gap of 100 bp
* 18431 38813: contig of 20383 bp in length
* 38814 38913: gap of 100 bp
* 38914 42991: contig of 4078 bp in length

```

```

FEATURES
source
* 42992 43091: gap of 100 bp
* 43092 46478: contig of 3387 bp in length
* 46479 46578: gap of 100 bp
* 46579 61624: contig of 15046 bp in length
* 61625 61724: gap of 100 bp
* 61725 70468: contig of 8744 bp in length
* 70469 70568: gap of 100 bp
* 70569 78278: contig of 7710 bp in length
* 78279 78378: gap of 100 bp
* 78379 80926: contig of 2648 bp in length
* 80927 81026: gap of 100 bp
* 81027 89376: contig of 8350 bp in length
* 89377 89476: gap of 100 bp
* 89477 92606: contig of 3130 bp in length
* 92607 92706: gap of 100 bp
* 92707 96333: contig of 3627 bp in length
* 96334 96433: gap of 100 bp
* 96434 114593: contig of 18160 bp in length
* 114594 114694: gap of 100 bp
* 114694 133054: contig of 18361 bp in length.
Location/Qualifiers
1..133054
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP5-852F22"
/clone_id="RPC1-5"
1..6858
/feature="assembly_fragment:01045"
/feature_chain:1
clone_end:SP6
vector_side:left"
6959..11344
/feature="assembly_fragment:00181"
/feature_chain:1"
11445..15274
/feature="assembly_fragment:01118"
/feature_chain:1"
15375..18330
/feature="assembly_fragment:00832"
/feature_chain:1"
18431..38813
/feature="assembly_fragment:00599"
/feature_chain:1"
38914..42991
/feature="assembly_fragment:00760"
/feature_chain:1"
43092..46478
/feature="assembly_fragment:00784"
/feature_chain:1"
46579..61624
/feature="assembly_fragment:00618"
/feature_chain:2"
61725..70468
/feature="assembly_fragment:01180"
/feature_chain:2"
70569..78278
/feature="assembly_fragment:00949"
/feature_chain:2"
78379..80926
/feature="assembly_fragment:00888"
/feature_chain:3"
81027..89376
/feature="assembly_fragment:00772"
/feature_chain:3"
89477..92606
/feature="assembly_fragment:00426"
/feature_chain:3"
92707..96333
/feature="assembly_fragment:00451"
/feature_chain:4"
96434..114593
/feature="assembly_fragment:00396"

```

```

misc_feature      fragment_chain:4"
114694..133054
/notes="assembly_fragment:00882
clone_end:T7
vector_side:right"
BASE COUNT      39306 a 27034 c 27151 g 38061 t 1502 others
ORIGIN
Query Match      1.1%; Score 21; DB 2; Length 133054;
Best Local Similarity 100.0%; Pred. NO. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1400 TTTTGAAGAAATTCAAATA 1420
|||||
Db 97814 TTTTGAAGAAATTCAAATA 97834

RESULT 22
AC114333/c
LOCUS            AC114333      145661 bp      DNA      linear      HTG 07-MAR-2002
DEFINITION      Canis familiaris clone RP81-588, WORKING DRAFT SEQUENCE, 4
unordered pieces.
AC114333
VERSION          AC114333.1 GI:19224982
KEYWORDS         HTG: HTGS_P1; HTGS_DRAFT.
SOURCE           Canis familiaris
ORGANISM         Canis familiaris
REFERENCE        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS         Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 145661)
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Maskeri,B.,
Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
Stantrop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Prasad,A.,
Taurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 145661)
Green,E.D.
Direct Submission
Submitted (07-MAR-2002) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nhgri.nih.gov
----- Project Information
Center project name: civ
Center clone name: 005B08
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 144074 bases at least Q40
Consensus quality: 144474 bases at least Q30
Consensus quality: 144791 bases at least Q20
Insert size: 120000; agarose-fp
Insert size: 145361; sum-of-contigs
Quality coverage: 9.62x in Q20 bases; agarose-fp
Quality coverage: 8.10x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

```

* be preserved.
* 1 6901: contig of 6901 bp in length
* 6902 7001: gap of unknown length
* 7002 45737: contig of 38736 bp in length
* 45738 45837: gap of unknown length
* 45838 90402: contig of 44565 bp in length
* 90403 90502: gap of unknown length
* 90503 145661: contig of 55159 bp in length.

FEATURES
source
1..145661
/organism="Canis familiaris"
/db_xref="taxon:9615"
/clone="RP81-588"
/clone_lib="RP81"
1..6901
/notes="assembly_fragment"
7002..45737
/notes="assembly_fragment"
45838..90402
/notes="assembly_fragment"
90503..145661
/notes="assembly_fragment"
BASE COUNT 43888 a 28373 c 28107 g 44993 t 300 others
ORIGIN

Query Match      1.1%; Score 21; DB 2; Length 145661;
Best Local Similarity 100.0%; Pred. NO. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1737 GTAAAGAAATCGCTCAAAAT 1757
|||||
Db 81634 GTAAAGAAATCGCTCAAAAT 81614

RESULT 23
AL356867
LOCUS            AL356867      147803 bp      DNA      linear      PRI 12-FEB-2002
DEFINITION      Human DNA sequence from clone RP1-20N18 on chromosome 1, complete
sequence.
ACCESSION        AL356867
VERSION          AL356867.23 GI:18655972
KEYWORDS         HTG.
SOURCE           human.
ORGANISM         Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 147803)
Ellington,A.
Direct Submission
Submitted (12-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 13, 2002 this sequence version replaced gi:16304699.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/c_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping

```

Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Ch1>  
 RPI-20N18 is from the library RPI-1 constructed by the group of  
 Pieter de Jong. For further details see  
<http://www.chori.crg/bacpac/home.htm>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone  
 RPI-20N18. It may be shorter because we sequence overlapping  
 sections only once, except for a short overlap.  
 The true right end of clone RPI-20N18 is at 147803 in this  
 sequence. The true left end of clone RPI-59H12 is at 107462 in this  
 sequence. The true right end of clone RPI-13P20 is at 2000 in this  
 sequence.

#### FEATURES

source

1. 147803  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RPI-20N18"  
 /clone\_lib="RPI-1"

BASE COUNT 45512 a 30941 c 29246 g 42104 t  
 ORIGIN

Query Match 1.1%; Score 21; DB 9; Length 147803;  
 Best Local Similarity 100.0%; Pred. No. 10; Indels 0; Gaps 0;  
 Matches 21; Conservative 0; Mismatches 0;

QY 39 CTGCTAATGCAATATTCG 59

Db 67488 CTGCTAATGCAATATTCG 67508

RESULT 24  
 AC011024/c

LOCUS 157326 bp DNA linear HTG 26-MAY-2000  
 DEFINITION Homo sapiens clone RPI-115D12, WORKING DRAFT SEQUENCE, 6 unordered  
 pieces.

AC011024

VERSION AC011024.3 GI:7107790

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

Web site: <http://www.seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: 11988

Center clone name: 115.D.12

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 142326 bases at least Q40

Consensus quality: 151063 bases at least Q30

Consensus quality: 154192 bases at least Q20

Insert size: 150000; agarose-fp

Insert size: 156826; sum-of-contigs

Quality coverage: 5.4 in Q20 bases; agarose-fp

Quality coverage: 5.2 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 8117: contig of 8117 bp in length  
 \* 8118 8217: gap of 100 bp  
 \* 8218 24311: contig of 16094 bp in length  
 \* 24312 24411: gap of 100 bp  
 \* 24412 46273: contig of 21862 bp in length  
 \* 46274 46373: gap of 100 bp  
 \* 46374 77849: contig of 31476 bp in length  
 \* 77850 77949: gap of 100 bp  
 \* 77950 112313: contig of 34364 bp in length  
 \* 112314 112413: gap of 100 bp  
 \* 112414 157326: contig of 44913 bp in length.

#### FEATURES

source

1. 157326  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RPI-115D12"  
 /clone\_lib="RPI-11 Human Male BAC"

misc\_feature

1. 8117  
 /note="assembly\_fragment"

misc\_feature

8218. 24311  
 /note="assembly\_fragment"

misc\_feature

24412. 46273  
 /note="assembly\_fragment"

misc\_feature

46374. 77849  
 /note="assembly\_fragment"

misc\_feature

77950. 112313  
 /note="assembly\_fragment"

misc\_feature

112414. 157326  
 /note="assembly\_fragment"

BASE COUNT

50401 a 28032 c 28903 g 49486 t 504 others

ORIGIN

Query Match 1.1%; Score 21; DB 2; Length 157326;

Best Local Similarity 100.0%; Pred. No. 10;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 TTTAAAAATCTGTGAAAG 557

Db 3618 TTTAAAAATCTGTGAAAG 3598

RESULT 25

AC106409

LOCUS

AC106409

168632 bp DNA linear HTG 17-JUL-2002

DEFINITION

Rattus norvegicus clone CH230-80K5, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
59 unordered pieces.

ACCESSION

AC106409

VERSION

AC106409.2 GI:21744154

KEYWORDS

HTG; HTGS PHASE1.

SOURCE

Norway rat.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 168632)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alabrooks,S.L., Amaraturunge,H.C., Ars,J.R., Ayele,M., Banks,T.,  
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsom,B., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,  
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Mosier,N., Neal,D., Newton,S., Newton,S., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokweso,S., Oguh,M., Okwuonu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,  
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.

Direct Submission  
Unpublished  
2 (bases 1 to 168632)  
Worley,K.C.  
Direct Submission  
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 168632)  
Worley,K.C.  
Direct Submission  
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 13, 2002 this sequence version replaced gi:18138930.

COMMENT

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
-----  
Center project name: GKVR

Center clone name: CH230-80K5  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; version 0.990329  
Assembly program: Phrap; version 0.990329  
Consensus quality: 119664 bases at least Q40  
Consensus quality: 125868 bases at least Q30  
Consensus quality: 130623 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 59 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 1501: contig of 1501 bp in length  
\* 1502: gap of unknown length  
\* 1602: contig of 1258 bp in length  
\* 2859: gap of unknown length  
\* 2959: gap of unknown length  
\* 4132: contig of 1193 bp in length  
\* 4252: gap of unknown length  
\* 5581: contig of 1229 bp in length  
\* 5481: gap of unknown length  
\* 7135: contig of 1554 bp in length  
\* 7235: gap of unknown length  
\* 8445: contig of 1210 bp in length  
\* 8545: gap of unknown length  
\* 9852: contig of 1307 bp in length  
\* 9952: gap of unknown length  
\* 1009: contig of 1009 bp in length  
\* 11061: gap of unknown length  
\* 12260: contig of 1199 bp in length  
\* 12360: gap of unknown length  
\* 13736: contig of 1376 bp in length  
\* 13836: gap of unknown length  
\* 14939: contig of 1103 bp in length  
\* 15039: gap of unknown length  
\* 16293: contig of 1254 bp in length  
\* 16393: gap of unknown length  
\* 17821: contig of 1428 bp in length  
\* 17921: gap of unknown length  
\* 19117: contig of 1196 bp in length  
\* 19217: gap of unknown length  
\* 20290: contig of 1073 bp in length  
\* 20390: gap of unknown length  
\* 22378: contig of 1988 bp in length  
\* 22478: gap of unknown length  
\* 24584: contig of 2106 bp in length  
\* 24684: gap of unknown length  
\* 26778: contig of 2094 bp in length  
\* 26878: gap of unknown length  
\* 28037: contig of 1159 bp in length  
\* 28137: gap of unknown length  
\* 29740: contig of 1603 bp in length  
\* 29840: gap of unknown length  
\* 31993: contig of 2153 bp in length  
\* 32093: gap of unknown length  
\* 34245: contig of 2152 bp in length  
\* 34345: gap of unknown length  
\* 35405: contig of 1060 bp in length  
\* 35505: gap of unknown length  
\* 37610: contig of 2105 bp in length  
\* 37710: gap of unknown length  
\* 38724: contig of 1014 bp in length  
\* 38824: gap of unknown length  
\* 40388: contig of 1564 bp in length  
\* 40488: gap of unknown length  
\* 42253: contig of 1765 bp in length  
\* 42353: gap of unknown length

```

* 42354 45557: contig of 3204 bp in length
* 42358 45557: gap of unknown length
* 45558 49442: contig of 3775 bp in length
* 49433 49532: gap of unknown length
* 49533 51958: contig of 2426 bp in length
* 51959 52059: gap of unknown length
* 52059 54469: contig of 2410 bp in length
* 54469 54569: gap of unknown length
* 54569 57443: contig of 2874 bp in length
* 57443 57542: gap of unknown length
* 57543 60504: contig of 2961 bp in length
* 60504 63230: gap of unknown length
* 63230 63321: contig of 2617 bp in length
* 63321 63321: gap of unknown length
* 63321 66198: contig of 2878 bp in length
* 66199 66298: gap of unknown length
* 66298 68570: contig of 2272 bp in length
* 68571 68571: gap of unknown length
* 68571 72180: contig of 3510 bp in length
* 72181 72280: gap of unknown length
* 72281 75540: contig of 3260 bp in length
* 75541 75640: gap of unknown length
* 75641 79577: contig of 3937 bp in length
* 79578 79677: gap of unknown length
* 79678 82755: contig of 3078 bp in length
* 82756 82855: gap of unknown length
* 82856 85669: contig of 2814 bp in length
* 85670 85769: gap of unknown length
* 85770 89449: contig of 3680 bp in length
* 89450 89549: gap of unknown length
* 89550 92645: contig of 3097 bp in length
* 92647 92746: gap of unknown length
* 92747 96376: contig of 3630 bp in length
* 96377 96476: gap of unknown length
* 96477 99555: contig of 3489 bp in length
* 99556 100055: gap of unknown length
* 100056 104630: contig of 4541 bp in length
* 104631 104706: gap of unknown length
* 104707 110232: contig of 5576 bp in length
* 110233 110233: gap of unknown length
* 110233 112931: contig of 2599 bp in length
* 112932 113031: gap of unknown length
* 113032 117078: contig of 3997 bp in length
* 117079 117178: gap of unknown length
* 117179 122292: contig of 5114 bp in length
* 122293 122392: gap of unknown length
* 122393 125710: contig of 3318 bp in length
* 125711 125810: gap of unknown length
* 125811 132052: contig of 6242 bp in length
* 132053 132152: gap of unknown length
* 132153 136836: contig of 4684 bp in length
* 136837 136937: gap of unknown length
* 136937 142183: contig of 5247 bp in length

```

Query Match 1.1%; Score 21; DB 2; Length 168632;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1034 ACATTGGATCTTCAGAAAG 1054  
 Db 114524 ACATTGGATCTTCAGAAAG 114544

RESULT 26  
 LOCUS HSA010770 170008 bp DNA linear PRI 07-JAN-1999  
 DEFINITION Homo sapiens hyperion gene, exons 1-50.  
 ACCESSION A010770  
 VERSION A010770.1 GI:4128007  
 KEYWORDS hyperion gene.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
 1 (bases 1 to 170008)  
 AUTHORS Kemmer, W. A.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-AUG-1998) Kemmer W.A., Department of Biochemistry,  
 Max-Planck Institute for Developmental Biology, Spemannstr. 35/11,  
 72076 Tuebingen, GERMANY  
 REFERENCE  
 2 (bases 1 to 170008)  
 AUTHORS Kemmer, W. A., Deiss, S. and Schwarz, U.  
 TITLE Cloning of Hyperion  
 JOURNAL Unpublished  
 FEATURES  
 source  
 1..170008  
 location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /map="q21-q22"  
 511..169570  
 /gene="hyperion"  
 join(511..558,29884..29919,33122..33379,39700..39744,  
 51569..51622,52296..52466,54032..54187,55014..55211,  
 60259..62646,71840..72053,73660..73739,75540..75678,  
 76428..76513,81649..81763,82225..82420,89306..89402,  
 90923..91015,97830..98183,100085..100309,101457..101597,  
 102079..102182,104419..104624,112137..112369,  
 120671..120833,121685..121897,124642..124874,  
 125838..125957,129441..129617,130316..130420,  
 13266..136418,137107..137286,138490..139563,  
 141933..142073,142581..143106,144230..144351,  
 144906..145097,145639..145827,148796..148940,  
 152508..152727,154434..154584,156100..156768,  
 156997..157205,157520..157625,159098..159280,  
 160257..160467,162005..162237,165089..165174,  
 166704..166833,167905..168044,169533..169570)  
 /gene="hyperion"  
 /evidence=experimental  
 join(511..558,29884..29919,33122..33379,39700..39744,  
 51569..51622,52296..52466,54032..54187,55014..55211,  
 60259..62646,71840..72053,73660..73739,75540..75678,  
 76428..76513,81649..81763,82225..82420,89306..89402,  
 90923..91015,97830..98183,100085..100309,101457..101597,  
 102079..102182,104419..104624,112137..112369,  
 120671..120833,121685..121897,124642..124874,  
 125838..125957,129441..129617,130316..130420,  
 13266..136418,137107..137286,138490..139563,  
 141933..142073,142581..143106,144230..144351,  
 144906..145097,145639..145827,148796..148940,  
 152508..152727,154434..154584,156100..156768,  
 156997..157205,157520..157625,159098..159280,  
 160257..160467,162005..162237,165089..165174,  
 166704..166833,167905..168044,169533..169570)  
 /gene="hyperion"  
 /codon\_start=1  
 /evidence=experimental  
 /product="Hyperion protein"  
 /protein\_id="CAA09361.1"  
 /db\_xref="GI:4128008"

/translation="MEDERQKLEANGKAKIEBELSLAFVROLAQRORAKSDGSP  
 SKOKKRRATSSSKHDVSAHDLNIDOSCENMYNSISGRVSTVIPESTIKRTLSHG  
 EITSHQGFVLESEIETPADDCSEVAGCSPVWRKPTMLLEEHEGVDSYSEQ  
 GAQSPHTLEMESELAGKQHEIEELNRELEEMRYVYGTGEGLOEFPAALIKORGI  
 ITQTLNANLOQARREKDETFMRLELITQESOKIOFOLOASBETLRNHSSTADLL  
 OKQOULTHQOOLEBODHLEDYOKKKEPTFOIISFOEKITRYEMEOCKYKNEKE  
 EIOEKETIIEELNTKIEBEKTLLEKDLTTADKLEGLQIOIVQKQETKOMKLEL  
 TNSKQERSSSEIKQMGTVELQKRNKDSQFEDIVQRMEOFTLEKQLEKRLD  
 EMYGOIIVQKQELIRQMAQMEEMKTRKSGEMENLRQYSINIVMEDIOKMMVAIN  
 EINTKLODINSQEKLEKELGLEKCALQOLEDLVELSEFSRQIOARQOTLAEQ  
 ESKNEAHKSLSTVEDIKAEIVSASERKELKHAETATYKIKLEMEKKNAYLD  
 RNASEQEALEKRLTQQLSHSEBSKLKEDIEIHRINIEKLNKIGTHYQOQDGL  
 QNMSQKIEKTFQEKQNLITKQNLILTSKLDQOQSLVNSKSEMTLQINELQKRI  
 EILROEKETKQLEQVQLOKTLLEKQMEKENDLOEKRAQLEAENSILNDEKRT  
 LEDMLKIHVPVQEEERLIFLDSIKSKSDSVKEKEIIEERNEDEKQOQILUNEIE  
 KQNTFSPAEKNFVNVYQELQEVACILKVKDLDLEDSKNKQLEYSKILANIELHL



ORINPTTVKMSVDFEDKTFVAETLMGEVVEKDDTTTLMKELSVTKREKLELSQRLS  
LDSQLQKHGEIIFLNEVSKLSQKEQVSKRCRELEIINHNRAENVQCDTQVSS  
LLDGVVMTSGAGSVSKVNSFGESKIMVEDKVSFENMTVGEESKEQELIIDHLP  
SVTSSSRATQPEKNDKELNKLSEQNDRLQMEARQICLSLTVSTHVDQVREY  
MENEKDKALCSKEBELPAQOEKIKELOKIHOLELQMTQBTGDEGKPLHILKQLO  
KAVSECSYFQTLCSVLGEVYTPALKCEVNAEDKENSQDYISENEDEPELODYEVQ  
DFOENHTLLANKYENKVLAVLOTRLSKIWQOQDCKMLFEGENLPKETEFLSIH  
SOMTNLEDDVNHKSKSSLODKLTKLEEQVQELSLISSLOQLKETEYQBAEIH  
CLQRKLQVSSSTVPPSLPVDVSVITSDSAQTMTYPGSCVKKNIDGTTFEFGFVKE  
ETNIVKLEKQYQLEEBEAVKVSIMSIAFAQTLSLISGKENTASSQAQHAVQ  
QOEYFNMKLSQDQIGQTFQTVDFVKFKEFKPLSKELGHEGKEIILSNSDPHDIPE  
SKDCLVITSEMFSKDTFIVROSITHDEIISVSSMDASQMLNEOLEDMFQELVROY  
QEOHQATELLQAQHWOMEROREQEOLOEIKRLNQLAQRSIDNENLVSERVL  
LEELALQLSIAGREKLCCLRSNSTQNGNENQGEVEQTFKEKELDKRDPDVP  
BILSNRVALQANNRLKILLEVVKTTAAVEETIGRHVLGILDRSSKQSQASLIWR  
SEASVKSCHVEHTRVDSIESPGSDMPNDINMWSKVTEEGTELSORLVRSFG  
AGTIDPNEELMNLISRLQAVKLEAISSETSSQLEHAKVTQTELMRESFQKOE  
ATESLKQOEERLRLHESRAREOLAVELSKAEGVIDGADKTLFERQIOEKTDIID  
RLEOBLLCASNRLOLEAEQOIOEERELLSROKEMKAEGPVQQLQOETKLMKE  
KLEVQCAKEVRDLOKQVKALEDVEQVSRFIELEOKNTELMDLRQQAQLEKQL  
EKMRFLDEQAIADREHNDVFQOETQLEQKLVVPRPOISEHOTREVEOLANHLKE  
KTDKSELLSKQEQDIOERNEIEKLEFRVRELEQALLVSADTFQKVDREHFGA  
VEAKPELSLEVQOAEARDAIDKEKEITNLEEQLOEPRELEKNENVEQOOLHOLEIO  
KKESTRLOELEQENKLPKDDMEKLGIAIKESDAMSDODQHLVFGKPAQIIQOEVEI  
DOLNEQVTKLQOQKITTDNKKVIBEKNELIRDLETOIECLMSDQECVKRREBIEQL  
NEVIEKQOELANTQKTSMAHLSREASLSKHQLOLVVIAKUALSQQVETANEWT  
FMKVLKETFNMQLTOEFLSKRERESVEKIQSIPENSVNVAIDHLSKDKPKLEVY  
LTEDALKSLENQTVFKFPEENGKSIINLETRLQLESTVSADKLELTQCYQKDMQ  
EQQOFETMLQKTVNFKQIVVEKVAALVSQIOLEAVQYAKFCQDNQTISSPERT  
NIONLNRLDELGSDISALTRLISELESQVEMHTSILILEKQVEIAEKNVLEKX  
LLEQLKLEGGKQREKKEKRSQDVVEVLKTTTELPHSNEESGFFNEALRAESVA  
TKALASVKEAKQLQELLVKTNTMTSLQKDLQVRDHLAAEKULSILEKDETEV  
QESKACMFELPKLSKSIASTQDGLTKISSNQTPQILVNAGIQINLQSECSSE  
VTETIISQFTEKIEHAAEILDMESRHSISETETLKREHYAVQLLKESCGTKAV  
IQCLRSKEVFGYNNCFSTLCDSGSDMGOGIYLTSHSQCFDIASBGRSEESASDSEP  
KKIKGLRAVNEGQVLSLTPESVSDGEDHSIOQVSEPMLEERKAVINTTSSIKOLI  
TKMOLQREAVYDSQSHESFDMRGELLLALQVFLBEERSVLLAAFRTELALGTITD  
AVGLNCLQEQIQSGVEIQAMCECLQKADRRSLSEIOALHAQNGKRIYLRQES  
AGSPOELLEYINQKQSOQMLEQVLSMDRATLQESKQVVAELKSELAQTK  
LESETTLKAQHKHLEAFRLVLEKTDVHLLNDTLASQKRSRELQWALEKEKAK  
LGRSEERKDELEDKFSLESQKORNLQNLLEQOQLLNEQSKIESQSMLEDAOL  
SEQGRNLEQVLVLESEKVRIREMSSTLDREHHAQLOQSSDGTGQSRPPLPSRDLK  
ELQOLKEKHSRIVELLNETEKYKLDLSLQTRQOKEKQVHRKTLQTEQANTBGCQK  
MHELOQVEDLQRLQEKQVQYLDLEQQLQIQMOEFQKQELEREBKRSRLIYQ  
NLNEFTTWSLSDTRNNVLQKIEGETKESNYAKLIEMNGGGTGCHHELEMIKQKQ  
CVASQLQVLPKASERIQFETADDEDFTWQENIDEIILQKLTGOQGEPSLVSPS  
TSCGSLTERLRQNAELTGHISQLTEKNDLRNMVKLEQIRWYROTGAGRDNSSRF  
SLNGANIEALIIASEKWNREKLTLOKSLKRAEAVYKLAELRNSLLOTLSPDES  
HVTLKRYGKLRASFRKALIYKKYLLLLGGFQCECDATLALLRMGQGPATDL  
EVTNRPKGFTFRSAVRVSIASRMKFLVRHVRVTGSVSNINRMDGFGLNQGAET  
DSFYHSSGGGLYGEPRHTYRSRDLTYRSLPFPQNRYPGTPADFPNPGSLACSQLQ  
NYDPDRALTDYITRLEALQRLGTTQSGSTTQFHAGMR

511. .558  
/gene="hyperion"  
/number=1

559. .29883  
/gene="hyperion"  
/number=1

29894. .29919  
/gene="hyperion"  
/number=2

29920. .33121  
/gene="hyperion"  
/number=2

33122. .33179  
/gene="hyperion"  
/number=3

33380. .39699  
/gene="hyperion"  
/number=3

39700. .39744  
/gene="hyperion"  
/number=4

exon  
intron  
exon  
intron  
exon  
intron  
exon  
intron  
exon

39745. .51568  
/gene="hyperion"  
/number=4

51569. .51622  
/gene="hyperion"  
/number=5

51623. .52295  
/gene="hyperion"  
/number=5

52296. .52466  
/gene="hyperion"  
/number=6

52467. .54031  
/gene="hyperion"  
/number=6

54032. .54187  
/gene="hyperion"  
/number=7

54188. .55013  
/gene="hyperion"  
/number=7

55014. .55211  
/gene="hyperion"  
/number=8

55212. .60258  
/gene="hyperion"  
/number=8

60112. .60258  
/gene="hyperion"  
/note="isoform 1-5A"

60259. .62646  
/gene="hyperion"  
/number=9

60525. .61129  
/gene="hyperion"  
/note="similarity to M.musculus EST C79026 (NID:92519356)"

60713. .61137  
/gene="hyperion"  
/note="similarity to M.musculus EST AAS42309 (NID:92288743) vk53c08.r1"

60736. .61049  
/gene="hyperion"  
/note="match to human EST AA044975 (NID:91523178) zfs3g11.r1"

61106. .61536  
/gene="hyperion"

Query Match 1.1%; Score 21; DB 9; Length 170008;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1395 ATTCTTTTTCGAAGAAATTC 1415  
|||||  
Db 95395 ATTCTTTTTCGAAGAAATTC 95415  
|||||

RESULT 27  
AC096466/c  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-84G6, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
60 unordered pieces.  
AC096466  
VERSION AC096466.3 GI:21723957  
KEYWORDS HTG; HTGS PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 175809)  
Muzny,D.M., Adams,C., Ali-osman,F.R., Allen,C.,  
Aisbrooks,S.L., Amarutunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,

exon  
intron  
exon  
intron  
exon  
intron  
exon

Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Earhart, D., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorelli, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
Homs, F., Howard, S., Huber, S., Hulky, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S.,  
Karissom, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsged, H.,  
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokweto, S., Ogun, M., Okunnu, G.,  
Oragunye, N., Oyedot, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rivers, M., Rojas, A., Rojudo, I., Rolfe, M., Ruiz, S., Savery, G.,  
Scherer, S., Scott, G., Shen, H., Shoochari, N., Sisson, I.,  
Sodergren, E., Sonike, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svarek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
Williams, G., Williamson, A., Wleczky, R., Woodson, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 175809)  
Worley, K.C.

Direct Submission  
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 175809)

Direct Submission  
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 10, 2002 this sequence version replaced gi:17956421.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: GFCO  
Center clone name: CH230-84G6

----- Summary Statistics  
Sequencing vector: Plasmid,  
Chemistry: Dye-terminator Big Dye: 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 122727 bases at least Q40  
Consensus quality: 126253 bases at least Q30  
Consensus quality: 129204 bases at least Q20

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see <http://www.hgsc.bcm.tmc.edu/docs/genbankdraftdata.html>).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 60 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1082: contig of 1082 bp in length  
1182: gap of unknown length  
2774: contig of 1592 bp in length  
2874: gap of unknown length  
4543: contig of 1669 bp in length  
4643: gap of unknown length  
5762: contig of 1119 bp in length  
5862: gap of unknown length  
6938: contig of 1076 bp in length  
7038: gap of unknown length  
8455: contig of 1417 bp in length  
8555: gap of unknown length  
9691: contig of 1136 bp in length  
9791: gap of unknown length  
10967: contig of 1176 bp in length  
11067: gap of unknown length  
12717: contig of 1650 bp in length  
12817: gap of unknown length  
14551: contig of 1734 bp in length  
14551: gap of unknown length  
16045: contig of 1394 bp in length  
16145: gap of unknown length  
16844: contig of 1539 bp in length  
17684: gap of unknown length  
18884: contig of 1100 bp in length  
18984: gap of unknown length  
20769: contig of 1785 bp in length  
20869: gap of unknown length  
22794: contig of 1925 bp in length  
22894: gap of unknown length  
24008: contig of 1114 bp in length  
24108: gap of unknown length  
25593: contig of 1485 bp in length  
25693: gap of unknown length  
27286: contig of 1593 bp in length  
27386: gap of unknown length  
28755: contig of 1369 bp in length  
28855: gap of unknown length  
30486: contig of 1631 bp in length  
30586: gap of unknown length  
32087: contig of 1622 bp in length  
32308: gap of unknown length  
33636: contig of 1328 bp in length  
33736: gap of unknown length  
34968: contig of 1232 bp in length  
35068: gap of unknown length  
36904: contig of 1836 bp in length  
37004: gap of unknown length  
39237: contig of 2233 bp in length  
39337: gap of unknown length  
40459: contig of 1122 bp in length  
40559: gap of unknown length  
42702: contig of 2143 bp in length  
42802: gap of unknown length  
46164: contig of 3362 bp in length  
46264: gap of unknown length  
48471: contig of 2207 bp in length  
48571: gap of unknown length  
51355: contig of 2784 bp in length  
51455: gap of unknown length  
53805: contig of 2350 bp in length  
53905: gap of unknown length  
55202: contig of 1237 bp in length  
55302: gap of unknown length  
57937: contig of 2635 bp in length  
58037: gap of unknown length  
61502: contig of 3465 bp in length  
61602: gap of unknown length

\* 61603 63399: contig of 1797 bp in length  
\* 63400 63499: gap of unknown length  
\* 63500 65079: contig of 3008 bp in length  
\* 66508 66607: gap of unknown length  
\* 66608 69630: contig of 3023 bp in length  
\* 69631 69730: gap of unknown length  
\* 69731 71599: contig of 1869 bp in length  
\* 71600 71699: gap of unknown length  
\* 71700 74129: contig of 2430 bp in length  
\* 74130 74229: gap of unknown length  
\* 74230 76644: contig of 2415 bp in length  
\* 76645 76744: gap of unknown length  
\* 76745 79598: contig of 2854 bp in length  
\* 79599 79698: gap of unknown length  
\* 79699 84161: contig of 4463 bp in length  
\* 84162 84261: gap of unknown length  
\* 84262 87349: contig of 3088 bp in length  
\* 87350 87449: gap of unknown length  
\* 87450 91094: contig of 3645 bp in length  
\* 91095 91194: gap of unknown length  
\* 91195 95118: contig of 3924 bp in length  
\* 95119 95218: gap of unknown length  
\* 95219 100153: contig of 4935 bp in length  
\* 100154 100253: gap of unknown length  
\* 100254 103585: contig of 3332 bp in length  
\* 103586 103685: gap of unknown length  
\* 103686 109212: contig of 5527 bp in length  
\* 109213 109312: gap of unknown length  
\* 109313 113568: contig of 4256 bp in length  
\* 113569 113668: gap of unknown length  
\* 113669 118480: contig of 4812 bp in length  
\* 118481 118580: gap of unknown length  
\* 118581 124124: contig of 5544 bp in length  
\* 124125 124224: gap of unknown length  
\* 124225 129675: contig of 5451 bp in length  
\* 129676 129775: gap of unknown length  
\* 129776 132662: contig of 5487 bp in length  
\* 132663 135362: gap of unknown length  
\* 135363 140677: contig of 5315 bp in length

Query Match 1.1% Score 21; DB 2; Length 175809;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1643 GTCCTTAATTTGCAATTTATT 1663  
|||||  
Db 86185 GTCCTTAATTTGCAATTTATT 86165

RESULT 28  
BSUB0011/c  
LOCUS  
DEFINITION Bacillus subtilis complete genome (section 11 of 21): from 2000171 to 207730  
Z99114 AL009126  
Z99114.1 GI:2634230  
KEYWORDS  
SOURCE  
ORGANISM  
Bacillus subtilis.  
Bacillus subtilis  
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
REFERENCE 1 (bases 1 to 207730)  
Kunst, F., Ogasawara, N., Moser, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borcher, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Conner, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Gollightly, E.J., Grandi, G., Guiseppe, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasehara, Y.,

Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parto, V., Pohl, T.M., Portetle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, P., Sekiguchi, J., Sekowska, A., Seror, S.J., Seror, P., Shin, B.S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzensegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yamamoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and Danchin, A.  
The complete genome sequence of the gram-positive bacterium Bacillus subtilis  
Nature 390 (6657), 249-256 (1997)  
98044033  
9384377  
PUBMED  
2 (bases 1 to 207730)  
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.  
Direct Submission  
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48  
FEATURES  
source  
1. 207730  
/organism="Bacillus subtilis"  
/strain="168"  
/db\_xref="taxon:1423"  
complement (10. 1436)  
/genes="yoeA"  
complement (10. 33)  
/genes="yoeA"  
complement (45. 1436)  
/genes="yoeA"  
/function="unknown"  
/notes="similar to hypothetical proteins from B. subtilis"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB13730.1"  
/db\_xref="GI:2634231"  
/db\_xref="SPTREMBL:O34474"  
/translation="METSQEPVLTQKSKPSVWRMSFLVPLLLSNVLSQGVGM  
MAVRWLGVDAVAASVFFPFLFLLISFTIGIGSSILIGQAVGAKNEERLKVCT  
TLTFTLLGVLAIVGSIETLDIRLMTGPENVIHVSANVARTLVAMPFELVYAT  
TFURGDSKTPYTLIVSTVINALLPVLILMGFKLGIYGSAYATVISTATFL  
VLMVLRKKHPLQFDKTRVRLKMDKELLVLLRLGVSPASINYLVSLEIAVISFV  
NHYGNATAAYGVNQVAVSQVPAVSLGIAVSFAAQSIGANEFRLKQVIRIATNR  
NYIIGVLIILIVFVSHQLSLFLTEQESLYIAHRLMTLWSYLFQAGIIISATNR  
ASGTVLWPTVISIFAIGVEVPVAFVLVSHVTKLEILGVWVGYPAFAVSLLVGYVQ  
FWKKKQITRLIQ"  
1722. 2345  
/genes="yoeB"  
1722. 2267  
/genes="yoeB"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB13731.1"  
/db\_xref="GI:2634232"  
/db\_xref="SPTREMBL:O34841"  
/translation="MKKCLLFTLTIALILSLSTNAPAKNTSGDLSQKQALQALSARE  
HFNTMSQNPVKVKAIVCSGTFEYQNLQVYMCSDLGTCAKAVNLTPLPTKATBK  
GFKDYHTVSKGLAVPIGDGNLLNKKSTAKLISKGGSTIITEFTVPTLDGSPSAK  
RKVTYFKNKKVKNQFDVVI"  
2274. 2292  
terminator

```

/gene="yosB"
319. .2315
/terminator
/gene="yosB"
2319. .2434
/gene="trnSL-Arg1"
2359. .2434
/gene="trnSL-Arg1"
2359. .2434
/product="trnSL-Arg1"
/complement(2745. .3032)
/gene="yocC"
/complement(2745. .3032)
/gene="yocC"
/function="unknown"
/codon_start=1
/tranl_table=11
/protein_id="CAB13732.1"
/db_xref="GI:2634233"
/db_xref="SPTREMBL:O35009"
/translation="MHIVQPIRELEKIOEVKQYILNKNRDYFLFIQINSALRLSDI
LPLOVDVCKNKHLMATESONEKEKDSIFRLAETGNINVERHEBKRLSF"
/complement(3327. .3578)
/gene="yocD"
/complement(3327. .3578)
/gene="yocD"
/complement(3348. .3578)
/gene="yocD"
/function="unknown"
/codon_start=1
/tranl_table=11
/protein_id="CAB13733.1"
/db_xref="GI:2634234"
/db_xref="SPTREMBL:O34555"
/translation="MYTIEIETAEITNLSEDIYKSLVBEKIRAVHDGEQFLYKQF
KTHLEQLENYKALVOETLINEPIEDIDVDED"
3763. .5826
/gene="ygcE"
3763. .5826
/gene="ygcE"
/EC_number="2.3.2.2"
/function="glutathione metabolism"
/note="alternate gene name: pac"
/codon_start=1
/tranl_table=11
/product="gamma-glutamyltranspeptidase"
/protein_id="CAB13734.1"
/db_xref="GI:2634235"
/db_xref="SWISS-PROT:P54422"
/translation="MKRTMVCLTALISVLIVAGSVPFHAEEKPKSYDEYKQVDVG
KQGVNATRAHPLASIGADV.KKGNATDAVAIOPLNTPRPMMSGIGGFMVYDVG
KITDVTIISRERAPACATPDMFLDENGKAIIPSEBVTGTAQVPGTGLBEALDK
WGRSMQKLTPTSTIKLEKGFIDSVLAELISDOEKLSTAKDVFUPNGEELKGD
TLIQDLAKTFKILRSKGTDAFYKGFAKTISDVODFGSMTEKEDLENDITIDEPI
WGDYGOIATTPPSSGGIFLLQMLKILHFNISQYDVRSEWKYQLAETMLSYDV
RASYAGDPEFNVNPLKGLHFDYIKERQQLINDQVKKPKAGDPMKYQSGSANYSQV
EQRKDKVSCOTHTTAVDWRGNVSYTTIEQLGEGINVPDYGVLINLETPEADP
GGANEVQPNKRPLUSSMTPTILFKDQKVLIVGSGAGTITSSVLOTILHEYEMELK
AAVEPRITNYSMSYREDEGVDPDVLKLMGMRKGTSPVDIGNVQISIDHENCT
FKGVADSSRNGAIGINLKRK"
/complement(5535. .6484)
/gene="yofA"
/complement(5535. .6484)
/gene="yofA"
/complement(5627. .6484)
/gene="yofA"
/function="unknown"
/note="similar to transcriptional regulator (LysR family)"
/codon_start=1
/tranl_table=11
/protein_id="CAB13735.1"
/db_xref="GI:2634236"
/db_xref="SPTREMBL:O34685"
/translation="WESGDLKIFQAVAREGSIKRAQMLNVQSNVTAARVNLLEDIN
IRLFRTRMRMKTIRAGENLIQYADQVLSLDQAEKSTRMSRQKXGPLRIGSLTMAV

```

```

THLPEHAASFLRRPEVDLSVNTADTHHLIQVLIDHKVDGAFVYQVEHAAROLHYS
HDEILVLSREGTAEDMIQQPMLFFGAGCSHRDKLLEBAGIHNKILIEFTLEAI
IKGVASAGGTLALPKSAVDGSEHRTNWIMHQLPPSYODLEIVFIYRDFITTSAFQTF
LDEINEMKR"
6612. .7601
/gene="yogA"
6612. .7601
/gene="yogA"
/function="unknown"
/note="similar to alcohol dehydrogenase"
/codon_start=1
/tranl_table=11
/protein_id="CAB13736.1"
/db_xref="GI:2634237"
/db_xref="SPTREMBL:O35017"
/translation="MKAVIHNGKAGLGLSVQDVPSTKPGYGEVVKLKSGALNRDL
FLMKRSQDPMHMLISGAGITIEIGGVKNVWYQFVVIPTLNDMLTENVPVPE
ILGGSDDITLAEVITPSQNAIKKPAIISMEAGVLPISATARAALPTKOLKGEH
LLIPIIGSVATVYALFMKAIQATVSVTSSEERKKALKGADYAFDYSNMBQLO
GKILDIVLDSIGPALFSEYFPHVKNRIVSFGASGDNLSPVRSLEFPQVNLVGTG
MGSGEFQAMLAFLDKHLRPVIDRIYPLEKACEAYKKMQEGRQFGNIGIYME"
/complement(7611. .9139)
/gene="glbB"
/complement(7611. .7625)
/gene="glbB"
/complement(7658. .9139)
/gene="glbB"
/EC_number="1.4.1.13"
/function="glutamate biosynthesis"
/codon_start=1
/tranl_table=11
/product="glutamate synthase (small subunit)"

```

```

Query Match 1.1% Score 21; DB 1; Length 207730;
Best Local Similarity 100.0% Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 749 CAACATACACAGTTGCATATG 769
Db 115484 CAACATACACAGTTGCATATG 115464

```

```

RESULT 29
AL732291 207890 bp DNA linear HTG 17-AUG-2002
LOCUS AL732291
DEFINITION Mus musculus chromosome X clone RP23-48114, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
ACCESSION AL732291
VERSION AL732291.6 GI:21615699
KEYWORDS HTG; HTGS PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 207890)
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Leonagomacromi, D.
JOURNAL Direct Submissions
Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jun 26, 2002 this sequence version replaced gi:21614687.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
Project Information
Center project name: BM48114
----- Summary Statistics
Assembly program: XGAP4, version 4.5
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator; 2% of reads
Chemistry: Dye-terminator Big Dye; 97% of reads

```

```

Consensus quality: 207379 bases at least Q40
Consensus quality: 207605 bases at least Q30
Consensus quality: 207722 bases at least Q20
Insert size: 207790; sum-of-contigs
Quality coverage: 9.87x in Q20 bases; sum-of-contigs Quality
coverage: 10.07x in Q20 bases; agarose-fp
*****
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 9063: contig of 9063 bp in length
*
* 9064 9163: Gap of 100 bp
*
* 9164 207890: contig of 198727 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..207890
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /chromosome="X"
                /clone="RP23-48I14"
                /clone_lib="RPCI-23"
            1..9063
                /note="assembly fragment:01288
                fragment chain:1"
            9164..207890
                /note="assembly fragment:04457
                fragment chain:1"
    misc_feature
        67167 a 38739 c 37935 g 63948 t 101 others
    misc_feature
        9164..207890
        /note="assembly fragment:04457
        fragment chain:1"
BASE COUNT 67167 a 38739 c 37935 g 63948 t 101 others
ORIGIN
Query Match
Best Local Similarity 1.1%; Score 21; DB 2; Length 207890;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 530 CAAAGTCTTAAAAATCTGTA 550
|||||
Db 119212 CAAAGTCTTAAAAATCTGTA 119232
|||||

RESULT 30
AC009630/c
LOCUS
DEFINITION Homo sapiens chromosome 8, clone RP11-360L9, complete sequence.
ACCESSION AC009630
VERSION AC009630.14 GI:19848415
KEYWORDS HTG.
SOURCE Homo sapiens
ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
    1 (bases 1 to 219622)
    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
    Baker,J., Baldwin,J., Barna,N., Beckler,R., Benn,J., Brown,A.,
    Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
    Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
    Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
    Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
    Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
    Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
    Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
    Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
    Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
    Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
    Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
    Tefaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
    Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
    Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
    Research, 320 Charles Street, Cambridge, MA 02141, USA
    3 (bases 1 to 219622)
    Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
    Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
    Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
    Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
    Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
    Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
    Gande,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
    Hagos,B., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C.,
    Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
    Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
    MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
    McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
    Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
    Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
    Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
    Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
    Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
    Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
    Strauss,N., Subramanian,A., Talamas,J., Tefaye,S., Theodore,J.,
    Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
    Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
    Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
    Direct Submission
    Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
    Research, 320 Charles Street, Cambridge, MA 02141, USA
    4 (bases 1 to 219622)
    Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
    Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
    Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
    Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
    Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
    Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
    Gande,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
    Hagos,B., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C.,
    Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
    Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
    MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
    McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
    Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
    Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
    Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
    Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
    Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
    Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
    Strauss,N., Subramanian,A., Talamas,J., Tefaye,S., Theodore,J.,
    Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
    Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
    Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
    Direct Submission
    Submitted (30-MAR-2002) Whitehead Institute/MIT Center for Genome
    Research, 320 Charles Street, Cambridge, MA 02141, USA
    On Mar 30, 2002 this sequence version replaced gi:19683283.
    All repeats were identified using RepeatMasker:
    Smit, A.F.A. & Green, P. (1996-1997)
    http://ftp.genome.washington.edu/RM/RepeatMasker.html
    ----- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
    ----- Project Information
    Center project name: L2093
    Center clone name: 360_L_9
    -----
    Location/Qualifiers
    1..219622
    /organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-360L9"
/clone_lib="RPC1-11 Human Male BAC"
complement(740..835)
/rpt_family="MIR"
repeat_region
1626..1730
/rpt_family="GA-rich"
repeat_region
1995..2019
/rpt_family="GC-rich"
2306..2368
/rpt_family="GA-rich"
complement(3606..3702)
/rpt_family="L2"
3987..4054
/rpt_family="MIR3"
5149..5504
/rpt_family="AluSx"
5900..6166
/rpt_family="AluSg"
6260..6432
/rpt_family="MIR"
6482..6363
/rpt_family="MLT1A1"
6877..7352
/rpt_family="MIR"
7370..7571
/rpt_family="AluSx"
7817..7914
/rpt_family="MERSA"
8016..8126
/rpt_family="MSTB1"
8117..8182
/rpt_family="MSTB1"
8183..8489
/rpt_family="AluSx"
8490..8555
/rpt_family="MSTB1"
9370..9577
/rpt_family="AluSx"
complement(9719..9832)
/rpt_family="MIR"
9928..10135
/rpt_family="MIR"
10239..10348
/rpt_family="L2"
10751..11044
/rpt_family="AluDb"
complement(11087..11322)
/rpt_family="L1MA7"
11324..11411
/rpt_family="AluY"
11412..11695
/rpt_family="AluSc"
complement(11696..12441)
/rpt_family="L1MA7"
12442..12744
/rpt_family="AluSx"
complement(12745..12803)
/rpt_family="L1MA7"
12804..13115
complement(12804..13115)
/rpt_family="AluSx"
complement(13116..13499)
/rpt_family="L1MA7"
13500..13792
/rpt_family="AluSx"
complement(13793..14061)
/rpt_family="L1MA7"
complement(14093..14239)
/rpt_family="MIR3"
14535..14557
/rpt_family="(GGGGA)n"

```

```

repeat_region complement(14642..14785)
/rpt_family="L1PA8"
repeat_region 15685..15702
/rpt_family="G-rich"
repeat_region complement(16451..16575)
/rpt_family="MER41A"
repeat_region 16576..16871
/rpt_family="AluSx"
complement(16872..17310)
/rpt_family="MER41A"
repeat_region 19693..19988
/rpt_family="AluDb"
repeat_region 19992..20023
/rpt_family="Alu"
repeat_region 20024..20288
/rpt_family="AluSg/x"
20289..20300
/rpt_family="Alu"
21314..22905
/rpt_family="SVA"
24485..24786
/rpt_family="AluSg"
complement(25042..25134)
/rpt_family="MIR3"
complement(26856..26938)
/rpt_family="MIR"
28565..28766

Query Match 1.1%; Score 21; DB 9; Length 219622;
Best Local Similarity 100.0%; Pred.No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 654 GGTTCGTTGACATTCATTT 674

Db 145761 GGTTCGTTGACATTCATTT 145741

```

RESULT 31
LOCUS AL365337 257967 bp DNA linear HTG 19-JUN-2002
DEFINITION Mus musculus chromosome 17 clone CT7-BM104010, *** SEQUENCING IN
PROGRESS ***, 39 unordered pieces.
ACCESSION AL365337
VERSION AL365337.5 GI:14575247
KEYWORDS HTG; HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 257967)
Sims,S.
Direct Submission
Submitted (17-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jun 28, 2001 this sequence version replaced gi:9994249.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM104010
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 234414 bases at least Q40
Consensus quality: 243875 bases at least Q30
Consensus quality: 249519 bases at least Q20
Insert size: 254167; sum-of-contigs
Insert size: 120873; 44.0% error; agarose-fp
Quality coverage: 3.08x in Q20 bases; sum-of-contigs Quality

```



```

misc_feature /note="assembly fragment:02033
fragment.chain:8"
118430..123121
/note="assembly fragment:00196
fragment.chain:8"
123222..136596
/note="assembly fragment:03035
fragment.chain:9"
136697..139702
/note="assembly fragment:00105
fragment.chain:9"
139803..146966
/note="assembly fragment:00045"
149067..151784
/note="assembly fragment:00250"
151885..154359
/note="assembly fragment:00546"
154460..156467
/note="assembly fragment:00581"
156568..160930
misc_feature
misc_feature
misc_feature

Query Match 1.1%; Score 21; DB 2; Length 257967;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1395 ATTCTTTTGAAGAAATTC 1415
Db 237441 ATTCTTTTGAAGAAATTC 237461

RESULT 32
AF322045 1472 bp DNA linear INV 02-MAY-2002
LOCUS Malawimonas jakobiformis chapteron subunit delta gene, partial
DEFINITION cds.
ACCESSION AF322045
VERSION AF322045.1 GI:18033033
KEYWORDS Malawimonas jakobiformis.
SOURCE Malawimonas jakobiformis.
ORGANISM Eukaryota; Malawimonadidae; Malawimonas.
REFERENCE 1 (bases 1 to 1472)
AUTHORS Archibald,J.M., O'Kelly,C.J. and Doollittle,W.F.
TITLE The chapteron genes of jakobid and jakobid-like flagellates:
implications for eukaryotic evolution
Mol. Biol. Evol. 19 (4), 422-431 (2002)
JOURNAL 21918420
MEDLINE 11919283
PUBMED 2 (bases 1 to 1472)
REFERENCE Archibald,J.M., O'Kelly,C.J. and Doollittle,W.F.
AUTHORS Archibald,J.M., O'Kelly,C.J. and Doollittle,W.F.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2000) Biochemistry and Molecular Biology,
Dalhousie University, 5859 College St., Halifax, NS B3H 4H7, Canada
FEATURES
Source
1..1472
Location/Qualifiers
1..1472
/organism="Malawimonas jakobiformis"
/strain="ATCC 50310"
/db_xref="ATCC:50310"
/db_xref="taxon:136089"
join(45..134,200..343,400..519,581..715,777..902,
965..1372,1446..1472)
/product="chapteron subunit delta"
join(45..134,200..343,400..519,581..715,777..902,
965..1372,1446..1472)
/note="CCRDelta; Tric"
/codon_start=1
/product="chapteron subunit delta"
/protein_id="AAU56961.1"
/db_xref="GI:18033034"
/translation="ITPTGDTVTINDGATILKQMDIQHPAGKMLVELSKSDIEAGD
GTSVVMAGSLTVOVQLLKGIHPISDSVLANKIGSETLQDMAFPVTLDRRO
LTKAVNTALSSKAHIFTSLSPADAVRVIDPATATNVLDKDIRVVRGLGATYDD
TELIDGLVLFDRPNEASAKRATGPKLVKAKIGLQPCLSAPRTNESHNLVVDGYDMQ

```

```

BASE COUNT 314 a 351 c 372 g 434 t 1 others
ORIGIN
Query Match 1.1%; Score 20; DB 3; Length 1472;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1709 GCACCACTCTGTGTGTGT 1728
Db 243 GCACCACTCTGTGTGTGT 262

RESULT 33
CAU38234 2895 bp DNA linear BCT 06-NOV-1996
LOCUS Clostridium acetobutylicum phosphotransacetylase (pta) and acetate
DEFINITION kinase (ack) genes, complete cds.
ACCESSION U38234
VERSION U38234.1 GI:1491788
KEYWORDS Clostridium acetobutylicum strain=ATCC 824.
SOURCE Clostridium acetobutylicum
ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 2895)
AUTHORS Boynton,Z.L., Bennett,G.N. and Rudolph,F.B.
TITLE Cloning, sequencing, and expression of genes encoding
phosphotransacetylase and acetate kinase from Clostridium
acetobutylicum ATCC 824
Appl. Environ. Microbiol. 62 (8), 2758-2766 (1996)
JOURNAL 96316392
MEDLINE 8702268
PUBMED 2 (bases 1 to 2895)
REFERENCE Boynton,Z.L., Bennett,G.N. and Rudolph,F.B.
AUTHORS Boynton,Z.L., Bennett,G.N. and Rudolph,F.B.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1995) Biochemistry and Cell Biology, Rice
University, 6100 South Main, Houston, TX 77005, USA
FEATURES
Source
1..2895
/organism="Clostridium acetobutylicum"
/strain="ATCC 824"
/db_xref="taxon:1488"
128..1129
/gene="pta"
128..1129
/gene="pta"
/note="catalyze the reaction from acetyl-coA to
acetyl-phosphate"
/codon_start=1
/product="phosphotransacetylase"
/protein_id="AA18300.1"
/db_xref="GI:1491789"
/translation="MDLIRSTWCKQKODKRIILABGEKRLIADKIKKGLAEV
LVGDENKIKERASELNLDISKAEIMDPETSLTEYARDFELRHKGWTFIKSEKLV
RPLVFAFMALKDGVDGVSQAVATGDLRLPGQIITAPGVKIVSGFVMIIPDC
DYGEGLLFAFDCAVNPNTSDLEADIAITTAETARKLCNVPKAMLSFTMGKAG
EMVDKKNVAVEITKKRPRLAIDGELQDAALDSRYALKAQSSNVAGNAVLPFDL
QTGNIGKLVQVFAPAKAIGPICQGFAPRINDLSRSCSESDLVNVVAITVQAQKI"
1141..2346
/gene="ack"
1141..2346
/gene="ack"
/note="catalyze the reaction from acetylphosphate to
acetate"
/codon_start=1
/translation="acetyl kinase"
/product="acetyl kinase"
/protein_id="AA18301.1"
/db_xref="GI:1491790"

```



```
/translation="MKNLVINGSSSIKYQFDMKDETVLAKGLVERIGIKGVITHK
VNGKIVTETPMEDHKRAIKULVDLALLNDEYGVIKNIIDEISAVGHRIVHGGEKYANSV
LIDEDVKISDCVSLAPLHNPPIIGINACKELMPNVPVAVFDTAFHQITPDYAYM
YAIPEYDYDKTKIKYGFHGTSHKYVSTAAEFICKVEDLKVVCHMGNGASITAVE
NGKSVDSMTGTPPGGLAMGARGSDMDPAVTFILMDKLININASVNNLNKSGIEGL
SGISSMEDIKKGNVVDKPKAMLAYSVFTVKIKQFISYTAVNGLDCLVFTGGIGE
NSFENREICKMNYLGIKIDDKQNDTMGIPMDISAGSKRVLVVPTNEELMIARD
TKDIVGKDK"
2422..->2895
/note="ORF X"
/codon_start=1
/translate=1
/product="unknown"
/protein_id="AAB18302.1"
/db_xref="GI:1491791"
/translation="MLDVSDLIITKKVKDKVDIQLGMYTFYDSEYKILEPLTKGK
LLMNSIITLVAEYGVIVELVCSCLKQFPYDIDLEDEKFTDMPENKDDREVFINNY
EVDINEIVENNIIISLPKIKLCREDCKGLCPMCGTNLISNCNCHDNIDPRFAKL"
BASE COUNT 1110 a 354 c 576 g 855 t
ORIGIN
Query Match 1.1%; Score 20; DB 1; Length 2895;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 480 ATGCAATGTTCTTGATTC 499
|||||
Db 927 ATGCAATGTTCTTGATTC 946
|||||
RESULT 34
IGU48714/c
LOCUS
DEFINITION Idiomylia grimshawi alcohol dehydrogenase (Adh) gene, complete cds.
ACCESSION U48714
VERSION U48714.1 GI:1256247
KEYWORDS
SOURCE Drosophila grimshawi.
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 3644)
Brennan,M.D., Thorpe,P.A., Hu,J. and Dickinson,W.J.
Molecular organization of the alcohol dehydrogenase loci of
Drosophila grimshawi and Drosophila hawaiiensis
Gene 181 (1-2), 51-55 (1996)
97128767
8973307
2 (bases 1 to 3644)
Brennan,M.D.
Direct Submission
Submitted (06-FEB-1996) Mark D. Brennan, Biochemistry, University
of Louisville, 319 Abraham Flexner Way, Louisville, KY 40292, USA
Location/Qualifiers
1. 3644
/organism="Drosophila grimshawi"
/strain="G1"
/db_xref="taxon:7222"
2116..3047
/gene="Adh"
join(2116..2208,2303..2707,2781..3047)
/codon_start=1
/product="alcohol dehydrogenase"
/protein_id="AAB40357.1"
/db_xref="GI:1256248"
/translation="MVJANGNIIFVAGLGGIGLTSREIVKSGFKNLVLDVNDPAA
IAELKLNPKVTITTFYFDVTVPLAETKLLKTFDKLKTVDLLINGAGILLDDNQIER
TIAVNTGTVNTTAAILDVFDKRGKGGVAVNICSVTGENSIYQVPVWSAKAALS
FTTSIAKLHITGTAVSINGPITKTVLHKFNSWLGVEPRVAELLEHPTQTTLQCA
QNFYKAEANGQAIKWLDRGLRAIEWTKHWDSGI"
BASE COUNT 1230 a 682 c 645 g 1087 t
ORIGIN
Query Match 1.1%; Score 20; DB 5; Length 7370;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 891 TTTTGGGAATGATTTTAAA 910
|||||
Db 3103 TTTTGGGAATGATTTTAAA 3084
|||||
RESULT 35
AF043383
LOCUS
DEFINITION Pleuronectes americanus aminopeptidase N (ampN) gene, partial cds.
ACCESSION AF043383
VERSION AF043383.1 GI:3452274
KEYWORDS
SOURCE Pseudopleuronectes americanus.
ORGANISM
Pseudopleuronectes americanus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
Pleuronectoidei; Pleuronectidae; Pseudopleuronectes.
1 (bases 1 to 7370)
Douglas,S.E., Gallant,J.W. and Bullerwell,C.E.
Molecular Investigation of Aminopeptidase N Expression in the
Winter Flounder, Pleuronectes americanus
J. Appl. Ichthyol. (1998) In press
2 (bases 1 to 7370)
Douglas,S.E., Gallant,J.W. and Bullerwell,C.E.
Direct Submission
Submitted (15-JAN-1998) Marine Biology, Institute for Marine
Biosciences, 1411 Oxford Street, Halifax, NS B0J 1G0, Canada
Location/Qualifiers
1. 7370
/organism="Pseudopleuronectes americanus"
/db_xref="taxon:8265"
<1658..>7370
/gene="ampN"
join(<1658..1988,2897..3203,3500..3642,3736..3872,
4142..4268,4483..4637,5129..5242,5364..5507,5744..5809,
6088..6153,6231..6403,6566..6642,6873..7006,7098..7153,
7361..>7370)
/gene="ampN"
/product="aminopeptidase N"
join(1658..1988,2897..3203,3500..3642,3736..3872,
4142..4268,4483..4637,5129..5242,5364..5507,5744..5809,
6088..6153,6231..6403,6566..6642,6873..7006,7098..7153,
7361..>7370)
/gene="ampN"
/EC number="3.4.11.2"
/codon_start=1
/product="aminopeptidase N"
/protein_id="AAC32807.1"
/db_xref="GI:3452275"
/translation="MGKGFYISKAVGVGVILGAGALATIIALSVVYSQEKAKNNNE
VSTVGPTGPTGPTTSPAVTTPAPSNAPKYPKPSLVPEHYNNKLPRLTADP
TTGLYFTGESTVEFKCTEDTLIHNSKNLTDELNGQWALSANVSGVAPAKT
SRLOPVTYLVLDQGLKLLQNYHLTFDTGELADDDGGFYRSVYVNEGQKVAVT
QMOTDARKAPFCDFEPAKCATFTNLIHDPETVALSNGAQRSPKVTIDGKLUKQTD
FEQTERKSTYLLAFIVSEFTSINNTVDNVLIRIPARKPAIDAGQAYALSKTGILKF
IAHELAMFGNLVTLRWNDLNLNEGFSASYVYLGADVAEPDWNIKDLIVLNDHVR
FAYDALASSHPLSAKEDIQRPQISELFPAISYSGASVIRMLSDFLTEDIFVWGUR
TYLKEFAFAGNAVYTDLWNHQLQMAVNTGTLPGSVQDMMTWLVMGFPVPTVINTTSG
EVSKQFLDLPDSEVTPSPFNKVIPIKTKRTATAQPPYVLEQSKATSKMTTGV
DWTLANLDVVGYRVNYDSDNWKLLKLVSTNHLQIVINRAQLVDDAFNLARAKI"
BASE COUNT 2197 a 1546 c 1483 g 2144 t
ORIGIN
Query Match 1.1%; Score 20; DB 5; Length 7370;
Best Local Similarity 100.0%; Pred. No. 40;
```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 743 AAGACCCACATACAGCTT 762  
Db 3328 AAGACCCACATACAGCTT 3347

RESULT 36  
AE001318/c 9989 bp DNA linear BCT 30-OCT-2000  
LOCUS AE001318 Chlamydia trachomatis section 45 of 87 of the complete genome.  
DEFINITION AE001318 AE001273  
ACCESSION AE001318.1 GI:3328875  
VERSION AE001318.1 GI:3328875  
KEYWORDS  
SOURCE Chlamydia trachomatis.  
ORGANISM Chlamydia trachomatis.  
REFERENCE 1 (bases 1 to 9989)  
AUTHORS Stephens,R.S., Kaiman,S., Lammell,C.J., Fan,J., Marathe,R.,  
Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,  
Koonin,E.V. and Davis,R.W.  
TITLE Genome sequence of an obligate intracellular pathogen of humans:  
Chlamydia trachomatis  
JOURNAL Science 282 (5389): 754-759 (1998)  
MEDLINE 9900809  
PUBMED 9784136  
REFERENCE 2 (bases 1 to 9989)  
AUTHORS Kaiman,S., Mitchell,W., Marathe,R., Lammell,C., Fan,J., Hyman,R.W.,  
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.  
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
JOURNAL Nat. Genet. 21 (4): 385-389 (1999)  
MEDLINE 9920606  
PUBMED 10192388  
REFERENCE 3 (bases 1 to 9989)  
AUTHORS Stephens,R.S., Kaiman,S., Lammell,C.J., Fan,J., Marathe,R.,  
Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,  
Koonin,E.V. and Davis,R.W.  
TITLE Direct Submission  
JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University  
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA  
FEATURES  
source  
1..9989  
/organism="Chlamydia trachomatis"  
/strain="D/UW-3/CX"  
/db\_xref="taxon:813"  
/complement (105..371)  
/gene="omcA"  
/note="CT444"  
complement (105..371)  
CDS  
/gene="omcA"  
/codon\_start=1  
/transl\_table=1  
/product="9kDa-Cysteine-Rich Lipoprotein"  
/protein\_id="AAC68043.1"  
/db\_xref="GI:3328876"  
/gene="omcA"  
/note="CT444"  
complement (105..371)  
CDS  
/gene="omcA"  
/codon\_start=1  
/transl\_table=1  
/product="hypotheetical protein"  
/protein\_id="AAC68048.1"  
/db\_xref="GI:3522897"  
/translation="MKRFLVATATALTGGCCSYPPKREYTLAOKCSPLFKHAQTL  
DFSLAIVFEMTFTKAPPCVSSLEELSON"  
complement (929..2449)  
/note="CT145"  
/gene="glcX"  
complement (929..2449)  
CDS  
/gene="glcX"

/codon\_start=1  
/transl\_table=1  
/product="Gutamy1-tRNA Synthetase"  
/protein\_id="AAC68044.1"  
/db\_xref="GI:3328877"  
/translation="MTQGNVVRVVAASPTGPDHVGATAYMALFNEVFAKNGOMILRI  
EDTQTRSDRYEANTFSAIKWCGIRMGDEGVGAAGPYQSETELYKKYAEILLO  
TDCAIKCATPQEOLEMRVAVASTGIRGQYRRIRYLSPEEVROQEBGQPTTILKV  
PLTGSVEVDQKGVFPVADVDQVIVKSDGFPTYIFANVVDHDLWGITHLGEF  
MISFTPKHLLEYAGWEPQEFHMLBLNDPGLSLSRKNPTSIFYRDAGYKEAF  
NMFLTMGSMGDEIEYMRQLEAPDPKIRGRGAFDILKIDMNMKHYLNHESGP  
ESLQELGWLINDEFLKILPICOSRLITTLADFVGLSPFTTALPOYSKEELPSL  
KOEQAVMLYGLVLYKLEKDIEMEDFPQGSKWLAARQVHHKRVITPLVATYGAK  
QGLPLFDSMELGKARTPARLTLYAQNLTGYSKKVQQQVDALQDPLEDRFLDF"  
complement (2721..3272)  
/gene="euc"  
/note="CT446"  
complement (2721..3272)  
CDS  
/gene="euc"  
/codon\_start=1  
/transl\_table=1  
/product="CHLPS Euc Protein"  
/protein\_id="AAC68045.1"  
/db\_xref="GI:3328878"  
/translation="MECLQDPTGVAREOVVOOQENNAVPYTSQSVSTQAKLHNT  
KQATYVAIKQKLTAKSTTRMTDLDQLEDIRRNRSRAKSTYQGLFLDNKGFYSV  
QAVSMALDVPQKITYATRIGAMKGERGSAVHVSEVDRYNDYLKKEERKQSL  
AAMREGFALGDLADAEHFPIS"  
complement (3677..5431)  
/gene="recJ"  
/note="CT447"  
complement (3677..5431)  
CDS  
/gene="recJ"  
/codon\_start=1  
/transl\_table=1  
/product="ssDNA Exonuclease"  
/protein\_id="AAC68046.1"  
/db\_xref="GI:3328879"  
/translation="MEKDSLSLGKMDYPRKNEVFLKILKEFLHAPALQVLTISRG  
FQSIIEIDFIPQLSLSHSTLSFLDMEKAVARLLQAAHNEHWIYDGDVDTGV  
TLIVFLOILGKTSYCCSGTLFPGHGTASLISMLDGLSLITVCGTAGAEVH  
AINKQIDIVITVDHMPGPKLPCHVAMNLPDKNRPYKNELTGTGVAPFKVCAEVE  
LIQDASMKDKDILRFIDVLSGITADVGLSGENRLIVSGYEIKAKRGKLGKLL  
CSLSGSDSEVSTNLGIRITPKNLSRLDSSQGVNLSHDPKNGITVSELSAV  
NOERQRIAEVLRDVERILANPKLTAGSAIVLASPNHRSRVTPIISARLARTYKPV  
AIIALQDGIKSGSLRTIGSPFLGVLRCESFTLSGCHDPAAGIMIKEDVGEFRKX  
FHLVSSSLRKDDAMRTLSLVGMDFSRINDLIASMLELPEFGNGVSPVYTAIO  
VRYPLTLAGNHVLYKLYNSGERNLBGTARCGQDKISLKAANNIPIDIVYTLIRMRSA  
RGATRLIQDPRIQIPRL"  
complement (5449..9751)  
/gene="secD/secF"  
/note="CT448"  
complement (5449..9751)  
CDS  
/gene="secD/secF"  
/codon\_start=1  
/transl\_table=1  
/product="Protein Export"  
/protein\_id="AAC68047.1"  
/db\_xref="GI:3328880"  
/translation="MKRNFKNLSIIILVPALALVHYVPTCLYSRPLNKRKIDGKAR  
KITERFTROVSDARNDLTLRVAKILSTLKDKIDKQSPISQVNVVHPLPDAALFTI  
DNVNGEISVPYKSAKRLVYVGSSEFTGTIVQISLVTSTLTKDLSVSYEDKQVTV  
SOELNLTNLNLFIESERPCEGYSILMSMPEIKVITYTRNISCGLRLPARKDTAF  
SRCLGTERDFTAFRLARLEKALOHBEVKAREBLKASVILHSRLNRNMSISPKVATV  
VDCSLSPSCFSMSLTADGKLLHFHDAVAANROQLVGNESLDLERLPAAMEKHATR  
LARSVEGTEPGCAVRLRDEAASGKIVLOGHVCQRIYVHLRLVYLNRIATGCDISTE  
NEPVYGERPLDALGCTFSPERSCRHFSKGSVITVKGSLSTVAKYEHGADEAAL  
LOHDLQNLVACFIHTDAVSWSLGEBRVLEIKEPDLRVIRWGESEVQSGFGASVLR  
VDRSLAVNRIEKTQHAELVRWDEQYRQAQCSMNRVYLRALIPKNIPLFEFLKKNIR  
KYSRGEHVLRFCTDIPVGGQKQIRIARDHQQILTDKAGIDKYSDELVYARLNGVSEV  
GMORGDHIOVSPGAAGISSADILGTSKMSGHVNVNROFSSRSPLRYVQTFDVLMP  
TARSLDECSPOAIRLAGALFRHNGNSAPANVAVAVETLRNGLSFSKLEBGSASLD

TOYSMTAIEKESREOVNPLMIVRNHLEGASLKNTRPEFAVGEVVLNFGVXDKATF  
SDRETPVQGFHVRHAKGGLFSGRGRMAVLDGVVIVPVLNVLK  
DHASVGNFYREHWAADLKGAMSFPEILSEEVVPELSSQVQVIGLWGL  
NVLIVFVYREYCGVTAIVLNLIIWASMOYLDAPLTLGLAGIILAMGMVDA  
NVLIVFEREYLLTSLSSVAGYKAFSAIPDSNLTTILASALLMLDGTPIKGF  
ALTIIIGIFSSMTALPMTKFFVFWQKTRQTLHMMKFIIGIKHNFLECKRLWV  
SGVVLGCVGLGAGWDSVGMDFKGYALTLDSDVCEYNPQMCVLRKFOQJL  
SSRDYRVRADSEKVIYLSQNALDRVEQIEGAGSEQSDHLARVLVLSDSGS  
TTSWVDFASGKSPKVSQJNSKWRTOQAVIALFGAIGIILFYVSLRPEWRYAFAICA  
LMHDLATCAVLVAFHQRIQIDLOAIGALMTVLGYSLNLTIIIPDRIRREKL  
FTWPIILINDALOKTGRVTWTTATLSVLVILLFVGGSGIFNFAFIMTVGILLGLTS  
SLVIAPLLLMVYKKEQNSLR"

BASE COUNT 3118 a 2399 c 1722 g 2750 t  
ORIGIN

Query Match 1.1%; Score 20; DB 1; Length 9989;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 761 TTGCATATGATGAATAAGTT 780  
|||||  
Db 6559 TTGCATATGATGAATAAGTT 6540

RESULT 37  
AB007683  
LOCUS  
DEFINITION Clostridium acetobutylicum ATCC824 section 171 of 356 of the complete genome.  
ACCESSION AB007683  
VERSION AB007683.1 GI:15024703  
KEYWORDS  
SOURCE Clostridium acetobutylicum  
ORGANISM Clostridium acetobutylicum  
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.  
REFERENCE 1 (bases 1 to 10742)  
Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q., Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I., Tatusov,R.L., Sabathe,P., Doucette-Stamm,L., Soucaille,P., Daly,M.J., Bennett G.N., Koonin,E.V. and Smith,D.R.  
Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum  
J. Bacteriol. 183 (16), 4823-4838 (2001)  
21359325  
11466286  
REFERENCE 2 (bases 1 to 10742)  
Childress,D., Zeng,Q. and Smith,D.R.  
Direct Submission  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100 Beaver Street, Waltham, MA 02453-8443, USA

FEATURES  
source  
1..10742  
/organism="Clostridium acetobutylicum"  
/strain="ATCC 824"  
/db\_xref="ATCC:824"  
/db\_xref="taxon:1488"  
111..2147  
/gene="CAC1736"  
/note="recG"  
111..2147  
/gene="CAC1736"  
/codon\_start=1  
/transl\_table=11  
/product="RecG helicase"  
/protein\_id="AAK79702.1"  
/db\_xref="GI:15024704"  
translation="MNINDDIATLKGVPKTLKLNKCGIFTILDLLLYPRDYENLS  
LMSNLEVEDQKVIKTPLRLEKFRSKRKTVTKIIIFSDGKTFSGIWFNOPYIK  
NKFYNTYTMGKIETKNEITWNPPIPVENRSSSEKIVPKVLSGTLKNTFFIRL  
IFELLQNIKIDENPFIKKNAFSLDESUREIHSFPRNMLESTRRLKPKQELFY  
SLKVLMLKEYIKKNNKGFAPKIAPELIKIKESLPFKUTEAQNULVIREILKDEKRPQAM

NRLVOGDSGKTIVALIAIPNVKNGFOAVLMAPTEILAKOHFESANTLKFKEFNIRI  
ELLTSVTDKNRIIKEKLDSEIDLIICTHALIEDNVFSNGLIIVVDELHRSFYMQ  
RRLFNKNNIDVLVMTATPIRTLALYLDGDDLIIDTLPGRKEVTKICLKNSR  
QKAYNPALNEIKNGRIQYVVCLEENKELETSVEKLYEELKSYFKGIKLEMLHKG  
MAPKENAEMERFNKTKETIVLSTTVIEGVNVPNASVMIIEDAQRFGLAQLHQLRGR  
VGRGOYQSCMLVAEMKNKVAERMEIMCSSSDGFIYSBEDFKLRGSGEVFGIRQSGE  
DGLILSDIIQDISILREANSAAKKLIISTEKSDIVIKKSIVKSLKNTSKYICPN"  
2225..2791  
/gene="CAC1737"  
2225..2791  
/gene="CAC1737"  
/codon\_start=1  
/transl\_table=11  
/product="N6-adenine-specific methylase"  
/protein\_id="AAK79703.1"  
/db\_xref="GI:15024705"  
translation="MRIISGKAKGRILPKPMGEMTRPTLDRVKEAMFNIONDVPEA  
VLDMFSGTSGLSGLEAASGAKVCYLIDKSPITYPILKENVENLRDEECKTNMDSY  
EAVYLSKSGKEFTLIFIDPPYAKEMIIPPAIDLICKEKLLTKSGLIVTKIDSDEKIYE  
GNNEIIVDHRKYKTIIVCFYRYKDKK"  
2788..3273  
/gene="CAC1738"  
/note="kdtB"  
2788..3273  
/gene="CAC1738"  
/codon\_start=1  
/transl\_table=11  
/product="Phosphopantetheine adenylyltransferase"  
/protein\_id="AAK79704.1"  
/db\_xref="GI:15024706"  
translation="MKVAVYVPGSFDPITNGHLDLIISRSKVFQKVIUGVLINPEKKCM  
FSVDREVDLIRVKIPFNNVVSQFSGLLVNFENNDNSVMIRGLRSVGDFEVLQTS  
LNNKLNPNVETVFMNTSUEFSFLSSTAIAQNAVFGCIKELVPDEIIDDLVLRKAEF  
N"  
3345..3869  
/gene="CAC1739"  
3345..3869  
/gene="CAC1739"  
/codon\_start=1  
/transl\_table=11  
/product="Uncharacterized conserved protein (coiled-coil)"  
/protein\_id="AAK79705.1"  
/db\_xref="GI:15024707"  
translation="MNVVKLLEYLGDIVDTSSKVPLSKNVKNVKNKEILDVITQMSEM  
PEELKQWILSEKDLNDVAKEANEIKQETETIKMQIDKHSITLEAEKRAKLLLE  
DAENEAKQIKLSALHYVDELLCHLDSOLDROQFNEKQVREASLSINVQDNFSIR  
EIIKNISRETVK"  
complement(3862..5007)  
/gene="CAC1740"  
complement(3862..5007)  
/gene="CAC1740"  
/codon\_start=1  
/transl\_table=11  
/product="Uncharacterized membrane protein, YLBj  
B.subtilis homolog"  
/protein\_id="AAK79706.1"  
/db\_xref="GI:15024708"  
translation="MIIFLLIIIFLITALLFLLIKELHLSNASNYIMTIFLSLIILF  
IILKPEVCISASISGAKLPFPVVPVPLIITNLIIICVDGIEIVSKLLGILCLPO  
RLPKKASIVLWVSLGYPGLGAKYAADLIENDIISFTFERLVNIASNPGLFIIGAV  
GTSMLKNKYLYILLISCYLSIANGIILPNKKDDFTKTKWNSKHSIGTAKTS  
TENAVKILQWMSFIIIFSVIIGILKQSVIFKNSSEAFIKFIILGIVEMTNGSSMLSS  
SSPIELKVTITISFLSSFGGLCVIAQIYSPYIGYKISYLFILKVIQGVISSIICFL  
LPKFLVYVNTISTFNANNFLSPSHFTVAIVLAVLPIIYKIKKTI"  
complement(5072..6280)  
/gene="CAC1741"  
complement(5072..6280)  
/gene="CAC1741"  
/codon\_start=1  
/transl\_table=11  
/product="Uncharacterized conserved protein, YLBM  
B.subtilis ortholog"  
/protein\_id="AAK79707.1"

Query Match	1.1%	Score 20;	DB 1;	Length 10742;
Best Local Similarity	100.0%	Pred. No. 39;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Db 7344	ATGCATGTTCTGTATTC	7363		
QY 480	ATGCATGTTCTGTATTC	499		
RESULT 38				
LOCUS	AE014001	10836 bp	DNA	linear BCT 26-JUL-2002
DEFINITION	Yersinia pestis KIM section 401 of 415 of the complete genome.			
ACCESSION	AE014001	AE009952		
VERSION	AE014001.1	GI:21960971		
KEYWORDS	.			
SOURCE	Yersinia pestis KIM.			
ORGANISM	Yersinia pestis KIM			
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Yersinia.			
AUTHORS	1 (bases 1 to 10836)			
	Deng, W., Burland, V., Plunkett, G., III, Boutin, A., Mayhew, G. F.,			
	Liss, P., Perna, N.T., Rose, D.J., Mau, B., Zhou, S., Schwartz, D.C.,			
	Fetherston, J.D., Lindler, L.E., Brubaker, R.R., Plana, G.V.,			
	Srereley, S.C., McDonough, K.A., Niles, M.L., Matson, J.S.,			
	Blattner, F.R. and Perry, R.D.			
TITLE	Genome Sequence of Yersinia pestis KIM			
JOURNAL	J. Bacteriol. 184 (16), 4601-4611 (2002)			
PubMed	12142430			
REFERENCE	2 (bases 1 to 10836)			
AUTHORS	Deng, W., Burland, V., Plunkett, G., III, Boutin, A., Mayhew, G. F.,			
	Liss, P., Perna, N.T., Rose, D.J., Mau, B., Zhou, S., Schwartz, D.C.,			
	Fetherston, J.D., Lindler, L.E., Brubaker, R.R., Plana, G.V.,			
	Srereley, S.C., McDonough, K.A., Niles, M.L., Matson, J.S.,			
	Blattner, F.R. and Perry, R.D.			
TITLE	Direct Submission			
JOURNAL	Submitted (21-PEB-2002) Genetics, University of Wisconsin, 445			
source	Henry Mall, Madison, WI 53706, USA			
FEATURES	Location/Qualifiers			
	1..10836			
	/organism="Yersinia pestis KIM"			
	/strain="KIM"			
	/db_xref="taxon:187410"			
	/complement (77..1066)			
	/gene="y3968"			
	/complement (77..1066)			
	/gene="y3968"			
	/function="unknown"			
	/note="residues 6 to 246 of 329 are 33.74 pct identical to			
	residues 7 to 248 of 329 from Genpept :			
	>gb AD37770.1 AF146532.10 (AF146532) glycosyltransferase			
	[Klebsiella pneumoniae]"			
	/codon_start=1			
	/transl_table=11			
	/product="hypothetical"			
	/protein_id="AA087512.1"			
	/db_xref="GI:21960972"			
	/translation="MNSVDISVIVPVNCANYLTLDLDSRSQTVSLIAVNDGSI			
	DHSLALEQVAHIDSRVLIINQOGSEARNVGIATAKGNWVAFAGDGDPAHALAS			
	TWLQYQKQKLDVILGNGFQSSNDEKPYPLIGQPMKATNGCOMITLSVQKEK			
	PIYAWDLIRRVISTHQLEPFRMLHEDILMTQQLVAQRISFCATPFYGYTNP			
	STNSPQALRANRANVYDINALVIAASTKRPVDALSHANSGHLGLRKK			
	LPSPQTAIAIRFITLGLPALFRGVTNLHEIMRAIRCSFLMGYSARQKAE"			
	complement (1201..2586)			
	/gene="y3969"			

```

CDS
complement(1201..2586)
/gene="y3969"
/function="putative transport"
/notes="residues 1 to 455 of 461 are 79.34 pct identical to
residues 180 to 633 of 637 from E. coli K12 : B3352;
residues 1 to 450 of 461 are 80.66 pct identical to
residues 180 to 629 of 635 from GenPept : >emb|CAD08154.1|
(AL627281) probable ABC transporter ATP-binding protein
[Salmonella enterica subsp. enterica serovar Typhi]"
/codon_start=1
/transl_table=11
/product="putative ATP-binding component of a transport
system"
/protein_id="AA087513.1"
/db_xref="GI:21960973"
/translation="MDLDAVWLKWLKSYTGLVLI SHDRDFLDPIIDKILHIEQQT
LNEYTGYSSFERQATKLSQQSQMYHQEKVAHLQSYIDRFRAQATKAAQASRIK
MLERMLIAPAHVDNPHFPTPELPPDLLRMDKVSAGYGVNTVLSQIKLNVPGS
RIGLLGRNGACKSTLIIKLACTLEPQSGEIALSKIGLYGAHQHLEFLRADEFLQHQ
MSRLAPKSEQLRDYLGCGYFGQDVTDFARTSGGKARLVLAIIHQRPNLILLD
EPTNHLIDLMKQALTEALIDFEGALVVVSHDRHLRSTDDLYLVHDKGVQPDGDL
DYQWLVDQRQQDQNPAPKELSGNSAQQRKDRDAEFRTQTQPLRKQIMTLEKQ
MDKLSLELAIAEQLADSALYDIARKADLTCLLQQTQVKSLEBETEMQWLDVQBLE
ILTKFESDADT"
complement(2547..3119)
/gene="y3970"
/complement(2547..3119)
/gene="y3970"
/function="transport"
/notes="N-terminal fragment of ATP-binding protein of ABC
transporter; naturally broken Orfs; residues 1 to 187 of
190 are 80.74 pct identical to residues 1 to 187 of 637
from GenPept : >gb|AA058460.1|AE005558.13 (AE005558)
putative ATP-binding component of a transport system
[Escherichia coli O157:H7 EDL933]"
/codon_start=1
/transl_table=11
/product="hypothetical"
/protein_id="AA087514.1"
/db_xref="GI:21960974"
/translation="MIVFSSLIQIRRGTRVLNDNATINPGQKVLGKNGCKGKSTLL
ALLKGLSADGNTVFPNNMALWNOETPALDIPATIEYVIDGDEYEQLEAELOAAN
EKNDGATATVHGKLDIAHANTIOSRAASLLHGLFGSQEKLQQPVRSPSGWRRLNL
AQALVCRSDLLLDLEPTTWIMWPSGKNG"
3367..3918
/gene="y3971"
3367..3918
/gene="y3971"
/function="putative"
/notes="residues 1 to 182 of 183 are 75.27 pct identical to
residues 1 to 182 of 184 from E. coli K12 : B3351"
/codon_start=1
/transl_table=11
/product="putative NAD(P)H oxidoreductase"
/protein_id="AA087515.1"
/db_xref="GI:21960975"
/translation="MMLQPPKVLVLYAHPEQSQSVANRVLQPVQQLHVTVDLYAH
YDPFDIHFHQQLIRHQVIVFQHPLYTSCPALLKEMLDRLVLRGANGVGGHALT
GKHWRSVITGQEQTIRIGGYNRYPMEDILRPFLTAAMCHMWINPMIYWARRQK
PETLASHAAQVQLMQLSTGL"
3922..5730
/gene="kefB"
/notes="kefB"
3922..5730
/gene="y3972"
3922..5730
/gene="kefB"
/function="transport; transport of small molecules;
cations"
/notes="residues 1 to 602 of 602 are 77.90 pct identical to
residues 1 to 601 from E. coli K12 : B3350"
/codon_start=1
/transl_table=11
/product="K+ efflux; NEM-activable K+/H+ antiporter"
/protein_id="AA087516.1"

```

---

```

/db_xref="GI:21960976"
/translation="MEGTGLLTAVLVFPAWVAVPIAQRIGIGAVGLYLIAGIAGP
WGLFIRDVDEILHSELVVFLFIIGLELNPALQWLRRSIFGVGAGQVVIITAAVL
IPALLYFTQFAWAAVIGGVGLAMSSTAMALQREKGNRNEGQLGFSVLLFODMAV
IPALALPILAGNEGANDWKIGLKTAAAPAGMLIGGRYLLRPLFRIVASGVREVPT
AAALLVLSALFMDALGSLMAGTFTIAGILLAESEFQHELEIAIEEPKGLLGLFFI
SVGNALDGLVFTLLDVLGVLFVIKSAIILGLARVGLRRSVLROFAGVLSQGG
EFAPVLSAAPSQRVLNAEQALLLVVVTLSMMTTPLLMQVIDRILLRVYNAQESDE
KPFVEDNPQVIVGFGFQGVGRLLMANKRITVLERDVSAYSMRKGYKYVYGD
ATELELLRAAEAKAIVITCNEPEDTMALVHLCCQHPFNLLILARARGVEAHELL
QNGVKDFTRETFSALBLGRKTLLELGMHPHOAVRAQHPRLDMRLRELMPQHGD
VQOISRIKEARRELEDFQREMLHESQLDGDWEYE"
5746..5967
/gene="y3973"
5746..5967
/gene="y3973"
/functions="unknown"
/notes="residues 11 to 73 of 73 are 65.07 pct identical to
residues 4 to 66 of 66 from GenPept : >gb|AAL22319.1|
(AE008858) putative cytoplasmic protein [Salmonella
typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="hypothetical"
/protein_id="AA087517.1"
/db_xref="GI:21960977"
/translation="MTITLNTKTRKRFIAGVCPQCNALDTLALWREDQVVEVCVK
CGHROQTDEQVKNKVRPHEQVIGIFPPE"
complement(5836..6063)
/gene="y3974"
complement(5836..6063)
/gene="y3974"
/functions="unknown"
/notes="residues 15 to 70 of 75 are 26.78 pct identical to
residues 112 to 167 of 1024 from GenPept :
>emb|CAA53232.1| (X75544) beta subunit RNA polymerase
[Plasmodium falciparum]"
/codon_start=1
/transl_table=11
/product="hypothetical"
/protein_id="AA087518.1"
/db_xref="GI:21960978"
/translation="MISPTLGTIGTDCSNLPRCTPHPKKSROIITLLWRKNTDHLML
WTNMFIDFLRLMVTTLNTHYFNLIPTP"
6062..6649
/gene="slyD"
6062..6649
/notes="slyD"
/gene="slyD"
/function="enzyme; proteins - translation and
modification"
/notes="residues 1 to 195 of 195 are 83.41 pct identical to
residues 1 to 196 of 196 from E. coli K12 : B3349;
residues 1 to 195 of 195 are 82.91 pct identical to
residues 1 to 196 of 196 from GenPept : >gb|AAL22318.1|
(AE008858) FKBP-type peptidyl prolyl cis-trans isomerase
(rotamase) [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="FKBP-type peptidyl-prolyl cis-trans isomerase
(rotamase)"
/protein_id="AA087519.1"
/db_xref="GI:21960979"
/translation="MKVTKDLVLSLAYQVRTEDGLVDESPVSLPYLHGHGSLIAG
LENALEHAGDSFDVRYNADEGYSDENLVQRPKDVPGVDELSVGRFLADTDQ
GPVPEITAVEDESHVVDGNHMLAGQDLNFHVEVAVRATGEBELQHGHHGEHDHHH
EHGDCGCGHGHDDHEHEHGGKCGKSGCGCH"
complement(6870..7088)
/gene="slyX"
/notes="slyX"
complement(6870..7088)
/gene="slyX"
/notes="residues 1 to 72 of 72 are 69.44 pct identical to

```

Query Match 1.1%; Score 20; DB 1; Length 10836;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 TGGCTGGCAATGCTTATG 323  
|||||  
Db 8567 TGGCTGGCAATGCTTATG 8566

RESULT 39  
AE002341/c 12442 bp DNA linear BCF 26-MAY-2000  
LOCUS Chlamydia muridarum, section 69 of 85 of the complete genome.  
DEFINITION AE002341 AE002160  
ACCESSION AE002341.2 GI:8163305  
VERSION  
KEYWORDS  
SOURCE Chlamydia muridarum.  
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
REFERENCE 1 (bases 1 to 12442)  
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,  
White,O., Hickey,H.K., Peterson,J., Umayam,L.A., Uetreback,T.,  
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,  
Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J.,  
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
Genome sequences of Chlamydia trachomatis Mohn and Chlamydia  
pneumoniae AR39  
TITLE Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
JOURNAL MEDLINE  
PUBMED 20150255  
10684935  
REFERENCE 2 (bases 1 to 12442)  
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,  
White,O., Hickey,H.K., Peterson,J., Umayam,L.A., Uetreback,T.,  
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,  
Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J.,  
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
Direct Submission  
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
On Jun 1, 2000 this sequence version replaced gi:7190754.  
COMMENT Location/Qualifiers  
FEATURES  
source  
1..12442  
/organism="Chlamydia muridarum"  
/db\_xref="taxon:83560"  
/note="s/nonym: Chlamydia trachomatis Mohn"  
complement (86..544)  
/gene="TC0726"  
complement (86..544)  
/gene="TC0726"  
/note="similar to SP:P18587, identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="15 kDa cysteine-rich outer membrane protein"  
/protein\_id="AAF39536.1"  
/db\_xref="GI:7190755"  
/translation="MSTPTVSGVTSQNNSSENVNNAASLTLLKERASKILSTARKV  
GLAVGIFVILISIVLFLIPATAENPYLAIIPALGCVNICIGILSNMKSGCSSEAK  
WKLCKNVLTQSEDLDDGELNNSNKIFTDNLSRVEDIYITLSSRSNSVA"  
complement (727..2391)  
/gene="TC0727"  
complement (727..2391)  
/note="similar to ; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="60 kDa outer membrane protein"  
/protein\_id="AAF39537.1"  
/db\_xref="GI:7190756"  
/translation="MRIGDPMNKLIRAVTIFAVTSVASLFGVLETMASSLSTNV  
ISLADTKAKETTSQKDRKARKKHQKRTSVKREVTAVADTKAVPRDSCGSKMTTV

KVNDNRNVEIVOSVEPATVSGPYPIETALGKRDQCVDTITQOLPCGEAFVSSDPAT  
TPTADGKLWKLIDPLRGCEKSKITWYWPLEKGCFTAAITACAGEITSYKCGPAT  
CVKQDGPESACLRCPVYTRINWNGTATKAVNENVPDGIATASQQRVUTYLG  
MOPGQRTITVECPCLKRGVNTATVATVGGHNTAVTIVNEPCVQVNI EGADWS  
YVCKEVEYIVSVSNPGDLVLRDVIEDTLSPGIVVEAAGAOISCNKLWTLKEINPG  
ESLQKVLVRAQOTPGQFTNNVYVKSQSCGCTSGTSAATYTWKGAATRMCVDPD  
ICVGENTYVYICVNRGSAEDNNVSLIKFSKELQPSFSGTGTITGNVYVFDLSL  
RLGKEIVESVYTLKAVASDARGESAILISDTLTVPSVDENTHY"  
complement (2539..2805)  
/gene="TC0728"  
complement (2539..2805)  
/note="similar to ; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="cysteine rich outer membrane protein 3"  
/protein\_id="AAF39538.1"  
/db\_xref="GI:7190757"  
/translation="MKRTALLAALCSVSLSSCCRIYDCCEPDPCAPIKSPCESKR  
EVNGGNSGNGVPSKCPGGBLDEHTKQGFQARQIQADGRCQ"  
3133..3357  
/gene="TC0729"  
3133..3357  
/gene="TC0729"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF39539.1"  
/db\_xref="GI:7190758"  
/translation="MKRFFLMITATALLTGCCSYEPKREYIIAQRCSPLFKIAQAI  
NFSLAQAVEKTEFTVTQPSLNLKEISQN"  
complement (3354..4874)  
/gene="TC0730"  
complement (3354..4874)  
/gene="TC0730"  
/note="similar to PID:1783380; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="glutamyl-tRNA synthetase"  
/protein\_id="AAF39540.1"  
/db\_xref="GI:7190759"  
/translation="WTFQNVVRVAPSPFGDPHVGTAYMALFENEYFARKFNGKMLIRI  
EDTQTRSIRDYEKNI FSAIKMCGIRMDGPDIGPYPYROSEETETIKYAEIILQ  
TDYAKCPATPOELOEMRAVASTLGVRGYRRARYLSAEVRQREOGQPTTLKTV  
PLTGESVPEDOCGRVVPMAVDVDQVLYKSDGPTVHFAVVDHLMGITHVLRGEE  
MLSTPRKULIYEARGWEPPOFFHHPLINDGSLSKRNKPTSTFYRIDAGYKKEAF  
NMFLLMGYSMEGDEILYSQRLIEAFPKIIGRGAVFDIRKLDWNNKHLYNHEGSP  
ESLKEILDMWLNDEFLIKLPLCQTRITTLADFLRLTSFFMALPEYSKDDLPSLSL  
GHQQAAMVLYSLVLYLEKDKDMEKDFYQSGSKMLAEAVQVHKKAVILPLVYITGK  
QGLPFDSEMLLGKARTARLTHAQNLLGVPKRLQOQIDRALQOQPLEIRPLDF"  
complement (5124..5681)  
/gene="TC0731"  
complement (5124..5681)  
/note="similar to GB:J13598 SP:Q06566 PID:289835;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="eno protein"  
/protein\_id="AAF39541.1"  
/db\_xref="GI:7190760"  
/translation="MECQODTMAVEVQVEOQEEENAVSANAQVITQAKLHV  
TQOATVYAIKOKKLKASKTTRMEITDLOLEBYRRNRYSRAKSTGELLFNEKGFYS  
VQVYASMLNVEQKIYATRIQAMKGERGSANWTHVSEVDRYRNDYIKAEARGKN  
LSAMKEGFTIGASELSDSEDFVS"  
complement (6143..7891)  
/gene="TC0732"  
complement (6143..7891)  
/gene="TC0732"

/note="similar to GB:MS4884 SP:P21893 PID:147550  
PID:897842 GB:U00096; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=1  
/product="single-stranded-DNA-specific exonuclease RecJ"  
/protein\_id="AAF39542.1"  
/db\_xref="GI:7190761"  
/translation="MEKDKLSALQKWTYPKHNEEFLLKKIKELYLHPVIAQLVSRG  
FQSVQEDRLYPOFSLHPASSFLDMKAVARLLKKENNEHVMYIGDGDVDGITGV  
TLVVEFGKQVTSYCCSGTLFKQHEITSLISQMLQDQVSLITVDCGFTAGREVR  
SINKQDVIWTDHTTGTGKLPKHCATLNPCLDKNSYPNKELTGVGAVKLCATYEE  
LIXODASQKIDLLHLHDVLVSLGTIADVGRSGENRILYSYGIKEITKGRGLKKL  
CSLSQVKSISISNLTGIRITPKNSLGRSLADSSQGVKLLSQDPRVTGIVSELAV  
NQERQAEVLRVIRILANPKLTAQSAIVLASPYWHSRVIPITISARLARTKPV  
AIALQDQISKDEATLTLPLDVCMDQFQISODLISSMELLEPPFGKGNPSPVYTRAQ  
FIHLVNSLRKDEATLTLPLDVCMDQFQISODLISSMELLEPPFGKGNPSPVYTRAQ  
VRYPLKSNHVKLYLNSGERNLGTAFGGQDKISLLKANNMELDIATLRYMKRSV  
HGAVRLLIQDPRIQAP"  
complement(8009. .12211)  
/gene="TC0733"  
complement(8009. .12211)  
/gene="TC0733"  
/note="similar to GB:AL009126; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="secDF protein, putative"  
/protein\_id="AAF39543.1"  
/db\_xref="GI:7190762"  
/translation="MKRNFRNLSVIFVAFALYHVLPTCLYYSRPLNKKIEGREAH  
KILRLTQQADARNIDTVRKRLSTKLKRGSVKPHLSIPGVNWHFPLAEDRAVFI  
DNVHGSHVPTKARLYLGSSELTGETITVQVSGSLVTLTEEDFSFVYEDKLVV  
SOELNATLFLNVVEGSCDHGYSLLDMSMPIEKVAHARGICSLGLLPTKTEOAPL  
SRFLGTQDFVAPLRVENALHEDVKEYEELNEAAVLLHSLRNWNEMS PRVGVNV  
VDCSELSPCFSSSLTGDGKLLHFDADVLAKQQLGDERLRLFAVEKQHIATR  
LNRSVETESGFAPLRDEAASGKIVLQHRVQCRIVEHLTALVNPRPIAETCDLSIE  
NFQVYREPMNDFITDGFIFSPSSCRHFSKGSYIVFKGLRSIIAKSCSSGQBEAGL  
LQDQLNLYACFTHTDVAVSLSGEDRVLEIKELQRIQVWGEFVQSGFEASLEVRD  
VRRLAVNRIEKAQHAELVRWDQYRQACSMNPQVRLRAAI PHKNVFFVFNKLVNR  
KYSRGEHLVRFQDFVCGKQIRLAFRDHOGNLTVDKEGIDKVSDELVARLNKLGVSF  
GMOREGHIQVSVPGANLSADILGTSKMSFHVNVNQFSRSPRLRYEVOTFDLYLWF  
TASRLSKSPGDNINHLAATFHEHQGFI PANVRAAIKGLAEAGLNFSEKHEMSACLD  
TQVSMIAVEGSGDNPLMVFNRHAGELASKNTRPEFAVGEVYLVNFGVDKAVS  
LDSKETPVQFHWTSKFCQGVNKTNSQFSGGRGWRMAVLDGYISDPVNLVPLK  
DHASVGNFVYREVRNLASDLKSGAMSFIPILSEEVVSPELGHSORLQGLSVFLGL  
VVLIALMSVYKFGGVIAVAILNALLIWASMOYLDAPLTLISLAGIILAMGWADA  
NVLVEERIREYLLSRSLSSVEAGYKAFSAIPDSNLTILASALLMLDGTGFKG  
ALTLIGIFSMTLSFTWTFPFLFIQKTRTQLDMMNKPFGVKNHFLKECKRLVWV  
SGAVVGLGIGLFGAWDSILGMDFGGYALTLDSDVCAYNPDMQSKALKRFRFGL  
SSRDYRVRKADSSGKVIYFSQNALARVEQAKGYSQNGADHHLAQVLOVLSDDSA  
TSPAAPFSSRSRGNFVKSQGLSNKRTQAIWALFGALGIIILLYSLRPEWRYAFSAICA  
LMHDLATCSVLVAFHFLHRIQIDLAQVGLMTVLGYSILNNTLIIIDRIEDRREKL  
FTWPIILLNDLQKLTGRVTWTTATLSLVILLFVGGSGIFNPFATMTVCGILLGLTUS  
SLIAPLPLLFMRKESQNSLQ"  
BASE COUNT 3968 a 2882 g 2074 g 3518 t  
ORIGIN  
Query Match 1.1%; Score 20; DB 1; Length 12442;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 647 ATTAAGGGTTTGCTTGAC 666  
|||||  
Db 9133 ATTAAGGGTTTGCTTGAC 9114  
|||||  
RESULT 40  
CEFS2H3/c 42632 bp DNA linear INV 24-JAN-2002  
LOCUS  
DEFINITION Caenorhabditis elegans cosmid F52H3, complete sequence.  
ACCESSION Z66512  
VERSION Z66512.1 GI:1041319

HTG.  
Caenorhabditis elegans.  
Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
1 none.  
Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
Science 282 (5396), 2012-2018 (1998)  
95069613  
9851916  
The C.elegans Sequencing Consortium.  
2 (bases 1 to 42632)  
Gardner,A.E.  
Direct Submissions  
Submitted (26-OCT-1995) Nematode Sequencing Project, Sanger  
Institute, Hinxton, Cambridge CB10 1SA, England and Department of  
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:  
jes@sanger.ac.uk or rwnematode.wustl.edu  
Coding sequences below are predicted from computer analysis, using  
predictions from Genefinder (P. Green, U. Washington), and other  
available information.  
Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.  
IMPORTANT: This sequence is not the entire insert of clone F52H3.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.  
The true left end of clone F52H3 is at 1 in this sequence. The true  
right end of clone F52H3 is at 1981 in  
sequence Z48543.  
The true left end of clone C18D1 is at 42533 in this sequence. The  
true right end of clone ZK892 is at 7399 in this sequence. The  
start of this sequence (1. .104) overlaps with the end of sequence  
Z48638.  
The end of this sequence (42533. .42632) overlaps with the start of  
sequence Z48543.  
For a graphical representation of this sequence and its analysis  
see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?  
name=F52H3  
IMPORTANT: This sequence is NOT necessarily the entire insert of  
the specified clone. It may be shorter because we only sequence  
overlapping sections once, or longer because we arrange for a small  
overlap between neighbouring submissions.  
Location/Qualifiers  
1. .42632  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="II"  
/clone="F52H3"  
join(Z48638.1:30587. .30627,Z48638.1:30651. .30750,  
Z48638.1:30795. .30954,Z48638.1:31010. .31190.105. .230,  
324. .405.449. .601)  
/gene="ZK892.5"  
/codon\_start=1  
/protein\_id="CAA91328.1"  
/db\_xref="GI:3877396"  
/db\_xref="SPTREMBL:Q23657"  
/translation="MRLVSFRAVTHISCOLKEYPEDATEIRELIRRRHYNVVSATA  
ILATFQWCTTVGVYIVLSAYAHMADSLVHISPMITTVFVLSVPSVIFPKKCAL  
LVPLTLLYFAAASLKLQFQIANQHLDFGRHFKSTTKYSEYGLLSLFDYSILVFTSI  
INCMCIGCSWHTFTPLVWNNFRKTEBAVALAKGLKTKKUSGNTITSSNDNSAHK  
LSHGAKGKRFISKHKQTKVQAQINENADVTTTSGAMDVKVGNQFLKLEIFWFOGE  
VRVA"  
complement(join(1113. .1226,1296. .1418,1464. .1625,



```

gene
    complement(join(17870..18124,18171..18458,18506..18784,
    18863..18990,19044..19083))
    /gene="F52H3.6"
    /complement(join(17870..18124,18171..18458,18506..18784,
    18863..18990,19044..19083))
    /gene="F52H3.6"
    /note="contains similarity to Pfam domain: PF00149
    (Ser/Thr protein phosphatase), Score=494.2,
    E-value=3.3e-145, N=1"
    /codon_start=1
    /protein_id="CAA91326.1"
    /db_xref="GI:3877394"
    /translation="MVNIQSFIDRLSVNDTSNPSTLRICNEADIDVIEKCTIILA
    QATVEVQAPLAVCGDIDHGOYTDLRIIPARGCFPPQONTPLGDVDRGOQLVLC
    LMAVIRYRNPREFIDRGNECASINTYGFYGEKRRYSLALVNDQNLFNHLPLCAM
    ISGRIFCHMGKISOKLWSMTOLAEITRPDPNNLSLAMDVLNADPENNTGMSSRG
    AVQIRGADVVDKPTERKMTDLARGQVVDGEEPPADMRIVTIFSPAPYCGFDNNA
    AMVIELRLVIFELIKPAIREVKIQARCAALAKMNSTSLVSSGRPKSKG"
    join(21426..21628,21681..21760,21812..21933,21978..221
    join(21426..21628,21681..21760,21812..21933,21978..221
    /gene="F52H3.5"
    /join(21426..21628,21681..21760,21812..21933,21978..221
    /gene="F52H3.5"
    /note="contains similarity to Pfam domain: PF00515 (TPR
    Domain), Score=60.9, E-value=9.1e-15, N=3
    cDNA EST yk573e3.5 comes from this gene"
    /codon_start=1
    /protein_id="CAA91325.1"
    /db_xref="GI:3877393"
    /db_xref="SPTREMBL:Q20683"
    /translation="WSTSHDRVAVLNQILNLPNFTSGIAEIHLESDSLERHVEYSL
    QLRESEVLAEGVRIDEAIKPTKALVECPKPSAVNNRQAQVRLQKPEKALDDNVE
    ALIAPKPTKTRICAOAVVORASITRLRGDDDKATDPASAEILOSSFPAKQVLNPPYA
    AMCNKMLAEVFEKAKGTGED"
    complement(join(22244..22462,22519..22958,23022..23425,
    23475..23858,23912..24182,24230..24428))
    /gene="F52H3.2"
    /complement(join(22244..22462,22519..22958,23022..23425,
    23475..23858,23912..24182,24230..24428))
    /gene="F52H3.2"
    /note="contains similarity to Pfam domain: PF01134
    (Glucose inhibited division protein A), Score=1405.3,
    E-value=0, N=1
    cDNA EST yk498a2.5 comes from this gene"
    /codon_start=1

Query Match      1.1%; Score 20; DB 3; Length 42632;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1397  TTCTTTTGAAGAAATTC 1416
Db 7940  TTCTTTTGAAGAAATTC 7921

RESULT 41
LOCUS      AC119990
DEFINITION Mus musculus clone RP24-506H15, LOW-PASS SEQUENCE SAMPLING.
ACCESSION  AC119990
VERSION    AC119990.1
KEYWORDS   HTG; HTGS PHASE0.
SOURCE     Mus musculus.
ORGANISM   Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 72371)
AUTHORS   Bitren,B., Linton,L., Nusbaum,C. and Lander,E.

```



TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Mus musculus, clone RP24-506H15  
Unpublished  
2. (bases 1 to 72371)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,  
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,  
Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kellis,C., LaRocque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Melidrim,T., Meneus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Roberti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Snit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE  
JOURNAL  
COMMENT

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L25778  
Center clone name: 506\_H\_15  
-----

\* NOTE: This record contains 88 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 723: contig of 723 bp in length  
\* 724 823: gap of 100 bp  
\* 824 1555: contig of 732 bp in length  
\* 1556 1655: gap of 100 bp  
\* 1656 2380: contig of 725 bp in length  
\* 2381 2480: gap of 100 bp  
\* 2481 3206: contig of 726 bp in length  
\* 3207 3306: gap of 100 bp  
\* 3307 4044: contig of 738 bp in length  
\* 4045 4144: gap of 100 bp  
\* 4145 4861: contig of 717 bp in length  
\* 4862 4961: gap of 100 bp  
\* 4962 5695: contig of 734 bp in length  
\* 5696 5795: gap of 100 bp  
\* 5796 6538: contig of 743 bp in length  
\* 6539 6638: gap of 100 bp  
\* 6639 7370: contig of 732 bp in length  
\* 7371 7470: gap of 100 bp  
\* 7471 8197: contig of 727 bp in length  
\* 8198 8297: gap of 100 bp  
\* 8298 9028: contig of 731 bp in length  
\* 9029 9128: gap of 100 bp

9129 9864: contig of 736 bp in length  
\* 9865 9964: gap of 100 bp  
\* 9965 10665: contig of 701 bp in length  
\* 10666 10765: gap of 100 bp  
\* 10766 11493: contig of 728 bp in length  
\* 11494 11593: gap of 100 bp  
\* 11594 12309: contig of 716 bp in length  
\* 12310 12409: gap of 100 bp  
\* 12410 13151: contig of 742 bp in length  
\* 13152 13251: gap of 100 bp  
\* 13252 13983: contig of 732 bp in length  
\* 13984 14083: gap of 100 bp  
\* 14084 14810: contig of 727 bp in length  
\* 14811 14910: gap of 100 bp  
\* 14911 15642: contig of 732 bp in length  
\* 15643 15742: gap of 100 bp  
\* 15743 16403: contig of 661 bp in length  
\* 16404 16503: gap of 100 bp  
\* 16504 17225: contig of 722 bp in length  
\* 17226 17325: gap of 100 bp  
\* 17326 18042: contig of 717 bp in length  
\* 18043 18142: gap of 100 bp  
\* 18143 18858: contig of 716 bp in length  
\* 18859 18958: gap of 100 bp  
\* 18959 19678: contig of 720 bp in length  
\* 19679 19778: gap of 100 bp  
\* 19779 20512: contig of 734 bp in length  
\* 20513 20612: gap of 100 bp  
\* 20613 21349: contig of 737 bp in length  
\* 21350 21449: gap of 100 bp  
\* 21450 22187: contig of 738 bp in length  
\* 22188 22287: gap of 100 bp  
\* 22288 23012: contig of 725 bp in length  
\* 23013 23112: gap of 100 bp  
\* 23113 23851: contig of 739 bp in length  
\* 23852 23951: gap of 100 bp  
\* 23952 24685: contig of 734 bp in length  
\* 24686 24785: gap of 100 bp  
\* 24786 25514: contig of 729 bp in length  
\* 25515 25614: gap of 100 bp  
\* 25615 26336: contig of 722 bp in length  
\* 26337 26436: gap of 100 bp  
\* 26437 27155: contig of 719 bp in length  
\* 27156 27255: gap of 100 bp  
\* 27256 27991: contig of 736 bp in length  
\* 27992 28091: gap of 100 bp  
\* 28092 28816: contig of 725 bp in length  
\* 28817 28916: gap of 100 bp  
\* 28917 29638: contig of 722 bp in length  
\* 29639 29738: gap of 100 bp  
\* 29739 30456: contig of 718 bp in length  
\* 30457 30556: gap of 100 bp  
\* 30557 31304: contig of 748 bp in length  
\* 31305 31404: gap of 100 bp  
\* 31405 32123: contig of 719 bp in length  
\* 32124 32223: gap of 100 bp  
\* 32224 32954: contig of 731 bp in length  
\* 32955 33054: gap of 100 bp  
\* 33055 33785: contig of 731 bp in length  
\* 33786 33885: gap of 100 bp  
\* 33886 34566: contig of 681 bp in length  
\* 34567 34666: gap of 100 bp  
\* 34667 35384: contig of 718 bp in length  
\* 35385 35484: gap of 100 bp  
\* 35485 36200: contig of 716 bp in length  
\* 36201 36300: gap of 100 bp  
\* 36301 37019: contig of 719 bp in length  
\* 37020 37119: gap of 100 bp  
\* 37120 37848: contig of 729 bp in length  
\* 37849 37948: gap of 100 bp  
\* 37949 38670: contig of 722 bp in length  
\* 38671 38770: gap of 100 bp  
\* 38771 39505: contig of 735 bp in length

```

* 39506 39605: gap of 100 bp
* 39606 40323: contig of 718 bp in length
* 40324 40423: gap of 100 bp
* 40424 41164: contig of 741 bp in length
* 41165 41264: gap of 100 bp
* 41265 41999: contig of 735 bp in length
* 42000 42099: gap of 100 bp
* 42100 42828: contig of 729 bp in length
* 42829 42928: gap of 100 bp
* 42929 43662: contig of 734 bp in length
* 43663 43762: gap of 100 bp
* 43763 44499: contig of 737 bp in length
* 44500 44599: gap of 100 bp
* 44600 45326: contig of 727 bp in length
* 45327 45426: gap of 100 bp
* 45427 46126: contig of 700 bp in length
* 46127 46226: gap of 100 bp
* 46227 46955: contig of 729 bp in length
* 46956 47055: gap of 100 bp
* 47056 47773: contig of 718 bp in length
* 47774 47873: gap of 100 bp
* 47874 48597: contig of 724 bp in length
* 48598 48697: gap of 100 bp
* 48698 49390: contig of 693 bp in length
* 49391 49490: gap of 100 bp
* 49491 50216: contig of 726 bp in length
* 50217 50316: gap of 100 bp
* 50317 51049: contig of 733 bp in length
* 51050 51149: gap of 100 bp
* 51150 51853: contig of 714 bp in length
* 51864 51963: gap of 100 bp
* 51964 52630: contig of 717 bp in length
* 52681 52780: gap of 100 bp
* 52781 53506: contig of 726 bp in length
* 53507 53606: gap of 100 bp
* 53607 54290: contig of 684 bp in length
* 54291 54390: gap of 100 bp
* 54391 55129: contig of 739 bp in length
* 55130 55229: gap of 100 bp
* 55230 55956: contig of 727 bp in length
* 55957 56056: gap of 100 bp
* 56057 56783: contig of 727 bp in length
* 56784 56883: gap of 100 bp

```

Query Match 1.1%; Score 20; DB 2; Length 72371;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 GAAAAATCAGCCAGTGTCT 152  
 Db 50904 GAAAAATCAGCCAGTGTCT 50885

RESULT 42  
 AC069119 73256 bp DNA linear HTG-09-MAY-2001  
 LOCUS Homo sapiens chromosome 17 clone CTD-2165M16 map 17, WORKING DRAFT  
 DEFINITION  
 AC069119 6 unordered pieces.  
 AC069119 3 GI:11597105  
 ACCESSION  
 AC069119.3 GI:11597105  
 VERSION  
 HTG, HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 KEYWORDS  
 Homo sapiens.  
 SOURCE  
 Homo sapiens  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 73256)  
 Birtten,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 17, clone CTD-2165M16  
 Unpublished  
 2 (bases 1 to 73256)  
 Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collimore,A., Cooke,P., Dearliano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kamm,L., Karakas,A.,  
 Klein,J., Labocque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGuirk,A., McKernan,K., McPheeters,R.,  
 Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zahoun,J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (18-MAY-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Dec 7, 2000 this sequence version replaced gi:9724653.  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

#### Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L8902

Center clone name: 2165 M.16

Summary Statistics

Sequencing vector: M13; M7815; 11% of reads

Sequencing vector: Plasmid; N/A; 89% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 72174 bases at least Q40

Consensus quality: 72665 bases at least Q30

Insert size: 63000; agarose-fp

Insert size: 72756; sum-of-contigs

Quality coverage: 5.9 in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently

consists of 6 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 1124: contig of 1124 bp in length

1125 1224: gap of 100 bp

1225 2355: contig of 1131 bp in length

2356 2455: gap of 100 bp

2456 5644: contig of 3189 bp in length

5645 5744: gap of 100 bp

5745 16114: contig of 10370 bp in length

16115 16214: gap of 100 bp

16215 23188: contig of 6974 bp in length

23189 23288: gap of 100 bp

23289 73256: contig of 49968 bp in length.

Location/Qualifiers

1. 73256

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="17"

/map="17"

/clone="CTD-2165M16"

/clone\_lib="CTD Human BAC"

1. 1124

#### FEATURES

source

misc\_feature

```
misc_feature /note="assembly_fragment"
1225..2355
misc_feature /note="assembly_fragment"
2456..5644
misc_feature /note="assembly_fragment"
5745..16114
misc_feature /note="assembly_fragment"
16215..23188
misc_feature /note="assembly_fragment"
23289..73256
BASE COUNT 23533 a 13938 c 13168 g 22117 t 500 others
ORIGIN
Query Match 1.1%; Score 20; DB 2; Length 73256;
Best Local Similarity 100.0%; Pred. NO. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1395 ATTTCCTTTTGAAGAAATT 1414
|||||
Db 17872 ATTTCCTTTTGAAGAAATT 17891

RESULT 43
AC087648/c
LOCUS AC087648 Homo sapiens chromosome 17 clone RP11-21M3 map 17, LOW-PASS HTG 15-JAN-2001
DEFINITION AC087648 Homo sapiens chromosome 17 clone RP11-21M3 map 17, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC087648
VERSION 1 GI:12229423
KEYWORDS HTG; HTGS PHASE0.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 74105)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP11-21M3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 74105)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Baatien,V., Boguslavskiy,L., Bouhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRoque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McSwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
Snit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
```

Center project name: L11999  
Center clone name: 21\_M\_3

\* NOTE: This record contains 88 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 731: contig of 731 bp in length  
\* 732 831: gap of 100 bp  
\* 832 1585: contig of 754 bp in length  
\* 1586 1685: gap of 100 bp  
\* 1686 2373: contig of 688 bp in length  
\* 2374 2473: gap of 100 bp  
\* 2474 3236: contig of 763 bp in length  
\* 3237 3336: gap of 100 bp  
\* 3337 4056: contig of 720 bp in length  
\* 4057 4156: gap of 100 bp  
\* 4157 4913: contig of 757 bp in length  
\* 4914 5013: gap of 100 bp  
\* 5014 5752: contig of 739 bp in length  
\* 5753 5852: gap of 100 bp  
\* 5853 6597: contig of 745 bp in length  
\* 6598 6697: gap of 100 bp  
\* 6698 7443: contig of 746 bp in length  
\* 7444 7543: gap of 100 bp  
\* 7544 8308: contig of 765 bp in length  
\* 8309 8408: gap of 100 bp  
\* 8409 9189: contig of 781 bp in length  
\* 9190 9289: gap of 100 bp  
\* 9290 10040: contig of 751 bp in length  
\* 10041 10140: gap of 100 bp  
\* 10141 10887: contig of 747 bp in length  
\* 10888 10987: gap of 100 bp  
\* 10988 11730: contig of 743 bp in length  
\* 11731 11830: gap of 100 bp  
\* 11831 12550: contig of 720 bp in length  
\* 12551 12650: gap of 100 bp  
\* 12651 13388: contig of 738 bp in length  
\* 13389 13488: gap of 100 bp  
\* 13489 14223: contig of 735 bp in length  
\* 14224 14323: gap of 100 bp  
\* 14324 15061: contig of 738 bp in length  
\* 15062 15161: gap of 100 bp  
\* 15162 15898: contig of 737 bp in length  
\* 15899 15998: gap of 100 bp  
\* 15999 16748: contig of 750 bp in length  
\* 16749 16848: gap of 100 bp  
\* 16849 17573: contig of 725 bp in length  
\* 17574 17673: gap of 100 bp  
\* 17674 18410: contig of 737 bp in length  
\* 18411 18510: gap of 100 bp  
\* 18511 19260: contig of 750 bp in length  
\* 19261 19360: gap of 100 bp  
\* 19361 20103: contig of 743 bp in length  
\* 20104 20203: gap of 100 bp  
\* 20204 20970: contig of 767 bp in length  
\* 20971 21070: gap of 100 bp  
\* 21071 21819: contig of 749 bp in length  
\* 21820 21919: gap of 100 bp  
\* 21920 22662: contig of 743 bp in length  
\* 22663 22762: gap of 100 bp  
\* 22763 23477: contig of 715 bp in length  
\* 23478 23577: gap of 100 bp  
\* 23578 24318: contig of 741 bp in length  
\* 24319 24418: gap of 100 bp  
\* 24419 25168: contig of 750 bp in length

```

* 25169 25268: gap of 100 bp
* 25269 26011: contig of 743 bp in length
* 26012 26111: gap of 100 bp
* 26112 26864: contig of 753 bp in length
* 26865 26964: gap of 100 bp
* 26965 27708: contig of 744 bp in length
* 27709 27808: gap of 100 bp
* 27809 28557: contig of 749 bp in length
* 28558 28657: gap of 100 bp
* 28658 29405: contig of 748 bp in length
* 29406 29505: gap of 100 bp
* 29506 30249: contig of 744 bp in length
* 30250 30349: gap of 100 bp
* 30350 31093: contig of 744 bp in length
* 31094 31193: gap of 100 bp
* 31194 31937: contig of 744 bp in length
* 31938 32037: gap of 100 bp
* 32038 32710: contig of 733 bp in length
* 32711 32870: gap of 100 bp
* 32871 33677: contig of 757 bp in length
* 33678 33727: gap of 100 bp
* 33728 34471: contig of 744 bp in length
* 34472 34571: gap of 100 bp
* 34572 35288: contig of 727 bp in length
* 35289 35398: gap of 100 bp
* 35399 36145: contig of 747 bp in length
* 36146 36245: gap of 100 bp
* 36246 36980: contig of 735 bp in length
* 36981 37080: gap of 100 bp
* 37081 37818: contig of 738 bp in length
* 37819 37918: gap of 100 bp
* 37919 38663: contig of 745 bp in length
* 38664 38763: gap of 100 bp
* 38764 39504: contig of 741 bp in length
* 39505 39604: gap of 100 bp
* 39605 40351: contig of 747 bp in length
* 40352 40451: gap of 100 bp
* 40452 41172: contig of 721 bp in length
* 41173 41272: gap of 100 bp
* 41273 42013: contig of 741 bp in length
* 42014 42113: gap of 100 bp
* 42114 42857: contig of 744 bp in length
* 42858 42957: gap of 100 bp
* 42958 43634: contig of 727 bp in length
* 43635 43784: gap of 100 bp
* 43785 44527: contig of 743 bp in length
* 44528 44627: gap of 100 bp
* 44628 45356: contig of 739 bp in length
* 45357 45466: gap of 100 bp
* 45467 46216: contig of 750 bp in length
* 46217 46316: gap of 100 bp
* 46317 47015: contig of 699 bp in length
* 47016 47115: gap of 100 bp
* 47116 47823: contig of 708 bp in length
* 47824 47923: gap of 100 bp
* 47924 48666: contig of 743 bp in length
* 48667 48766: gap of 100 bp
* 48767 49506: contig of 740 bp in length
* 49507 49606: gap of 100 bp
* 49607 50350: contig of 744 bp in length
* 50351 50450: gap of 100 bp
* 50451 51179: contig of 729 bp in length
* 51180 51279: gap of 100 bp
* 51280 52021: contig of 742 bp in length
* 52022 52121: gap of 100 bp
* 52122 52850: contig of 729 bp in length
* 52851 52950: gap of 100 bp
* 52951 53684: contig of 734 bp in length
* 53685 53784: gap of 100 bp
* 53785 54535: contig of 751 bp in length
* 54536 54635: gap of 100 bp
* 54636 55392: contig of 757 bp in length
* 55393 55492: gap of 100 bp

```

```

* 55493 56262: contig of 770 bp in length
* 56263 56362: gap of 100 bp
* 56363 57105: contig of 743 bp in length
* 57106 57205: gap of 100 bp
* 57206 57966: contig of 761 bp in length
* 57967 58066: gap of 100 bp

Query Match      1.1%; Score 20; DB 2; Length 74105;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1410 AATTCGAATGATTGGCA 1429
Db 48377 AATTCGAATGATTGGCA 48358
|||||

```

```

RESULT 44
AL713969      84574 bp  DNA    linear  PRI 20-JUN-2002
LOCUS
DEFINITION
Human DNA sequence from clone Rpl1-501117 on chromosome 10,
complete sequence.
ACCESSION
AL713969
VERSION
AL713969.11  GI:21537520
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 84574)
Heath,P.
Direct Submission
Submitted (31-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humney@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jun 21, 2002 this sequence version replaced gi:21212454.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; SW:
SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
Rpl1-501117 is from the library RPl-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

```

#### FEATURES

source

```

1..84574
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="Rpl1-501117"
/clone_lib="RPl-11.2"

```

#### BASE COUNT

```

27476 a 16456 c 15040 g 25602 t

```

```

Query Match      1.1%; Score 20; DB 9; Length 84574;
Best Local Similarity 100.0%; Pred. No. 38;

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 17, 2002, 12:32:52 ; Search time 17.5 Seconds  
(without alignments)  
6267.923 Million cell updates/sec

Title: US-09-868-987-1  
Perfect score: 3312  
Sequence: 1 atggaattccgcatattgtc.....atgacaattcagataatgc 1864

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Xgapop 10.0	Xgapext 0.5
Xgapop 6.0	Xgapext 7.0
Delop 6.0	Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09868987/runat 16122002 083007 23590/app query.fasta\_1.2055  
-DB=Issued Patents AA -QPMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOEPCI=0  
-LOEPCXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=45  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09868987 @CGN 1 1 0 @runat 16122002 083007 23590 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGESUBRY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WEN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6 -FCGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pcp.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pcp.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pcp.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pcp.\*
- 5: /cgn2\_6/ptodata/1/iaa/PCUTS.COMB.pcp.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	678.5	20.5	737	4	US-09-462-844-2
2	647.5	19.6	766	4	US-09-134-001C-3689
3	364	11.0	323	4	US-09-462-844-4
4	359.5	10.9	615	4	US-09-462-844-3
5	136.5	4.1	1051	4	US-09-134-001C-5005
6	117	3.5	474	3	US-09-022-699-2
7	108	3.3	901	4	US-09-134-001C-5389
8	103.5	3.1	383	1	US-08-196-989B-4
9	103.5	3.1	383	2	US-08-760-936-4
10	101	3.0	381	2	US-08-845-566-3
11	101	3.0	381	2	US-08-467-948A-28
12	101	3.0	381	3	US-08-852-824-18

13	101	3.0	381	3	US-08-467-947A-28	Sequence 28, Appl
14	101	3.0	381	5	PCT-US96-10618-4	Sequence 4, Appl
15	99.5	3.0	524	4	US-09-134-001C-3353	Sequence 3353, Ap
16	97.5	2.9	382	4	US-09-462-733-2	Sequence 2, Appl
17	97.5	2.9	1820	3	US-07-998-289B-8	Sequence 8, Appl
18	97.5	2.9	2100	2	US-08-808-793-23	Sequence 23, Appl
19	97.5	2.9	2100	3	US-08-772-512A-19	Sequence 19, Appl
20	97	2.9	400	4	US-09-134-001C-2912	Sequence 2912, Ap
21	97	2.9	1090	3	US-08-307-896-3	Sequence 3, Appl
22	97	2.9	1090	3	US-08-726-214-4	Sequence 4, Appl
23	97	2.9	1090	5	PCT-US95-11808-3	Sequence 3, Appl
24	96.5	2.9	382	4	US-09-262-477-2	Sequence 2, Appl
25	95	2.9	518	4	US-09-134-001C-4744	Sequence 4744, Ap
26	94	2.8	352	1	US-08-196-989B-2	Sequence 2, Appl
27	94	2.8	352	2	US-08-760-936-2	Sequence 2, Appl
28	93.5	2.8	490	4	US-09-134-001C-5116	Sequence 5116, Ap
29	93.5	2.8	1285	2	US-08-540-406-6	Sequence 6, Appl
30	93.5	2.8	1285	3	US-08-656-055-6	Sequence 6, Appl
31	93.5	2.8	1285	4	US-08-954-668-6	Sequence 6, Appl
32	93.5	2.8	1285	4	US-08-918-658-6	Sequence 6, Appl
33	93.5	2.8	1285	5	PCT-US95-13233-6	Sequence 6, Appl
34	93.5	2.8	1286	4	US-09-268-140-3	Sequence 3, Appl
35	93.5	2.8	1299	4	US-08-460-900C-62	Sequence 62, Appl
36	93.5	2.8	1299	4	US-08-674-509B-48	Sequence 48, Appl
37	93.5	2.8	1299	4	US-08-954-698-48	Sequence 48, Appl
38	93	2.8	344	4	US-09-134-001C-5567	Sequence 5567, Ap
39	92.5	2.8	334	1	US-08-118-270-73	Sequence 73, Appl
40	92.5	2.8	334	5	PCT-US93-08528-73	Sequence 73, Appl
41	92	2.8	335	4	US-09-387-699-2	Sequence 2, Appl
42	92	2.8	335	4	US-09-641-259B-2	Sequence 2, Appl
43	91	2.7	269	1	US-08-118-270-64	Sequence 64, Appl
44	91	2.7	269	5	PCT-US93-08528-64	Sequence 64, Appl
45	91	2.7	341	2	US-08-846-762-92	Sequence 92, Appl

#### ALIGNMENTS

RESULT 1  
US-09-462-844-2  
; Sequence 2, Application US/09462844  
; Patent No. 6258563  
; GENERAL INFORMATION:  
; APPLICANT: Quax, Wilhelmus J.  
; TITLE OF INVENTION: Increasing Production of Proteins in Gram-Positive Microorganisms  
; FILE REFERENCE: GC385-US  
; CURRENT APPLICATION NUMBER: US/09/462,844  
; CURRENT FILING DATE: 2000-01-13  
; PRIOR APPLICATION NUMBER: PCT/US98/14786  
; PRIOR FILING DATE: 1998-07-15  
; PRIOR APPLICATION NUMBER: EP 97305286.3  
; PRIOR FILING DATE: 1997-07-16  
; PRIOR APPLICATION NUMBER: EP 97305344.0  
; PRIOR FILING DATE: 1997-07-17  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 737  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-462-844-2

Alignment Scores:  
Pred. No.: 3.82e-69 Length: 737  
Score: 678.50 Matches: 169  
Percent Similarity: 50.43% Conservative: 125  
Best Local Similarity: 28.99% Mismatches: 192  
Query Match: 20.49% Indels: 97  
DB: 4 Gaps: 15

US-09-868-987-1 (1-1864) x US-09-462-844-2 (1-737)

Qy 104 GTCAGCAGCCCTATTTTAAACGTCCTCCATTGAAA---AATCATGCCAGTGTCTCAGGAAA 160

```

Db 197 ValSerIaProaenValSerIngluLeuAsnThrThrAspValysIleGluGlyHis 216
Qy 161 TTAAACCAACCGTGAAGTGAAGCAAACTGCGCTCAGATTAAATGGAGCGATCTTTT 220
Db 217 PheThrAlaGlnGlnAlaIaIaYAspLeuAlaSerIleLeuAlaAlaGlyAlaLeu 234
Qy 221 GTTCCCAAGGTTCTCAGTGAAGAG-----ACGATCTCTGTATCTTTGGAAAAA 271
Db 235 ---ProValysLeuThrGlnIleYrSerThrSerValGlyAlaGlnPheGlyGln 253
Qy 272 CAATGTACACAAAGCATTATCTCAGACAGTCGTGCGTGGCAATGCTTATGTTGATG 331
Db 254 AlaLeuHisAspThrValPheAlaGlyIleValGlyIleAlaIleIlePheLeuPheMet 273
Qy 332 AGCGTATATATATAGATTGAGGCGTCATCGCGAGCTGTCTTGTGAATCTTTTG 391
Db 274 LeuPheTyrrYrArgLeuPProGlyLeuIle-----AlaValIleThrLeuSerVal 290
Qy 392 CTATCTGGGACGCTTACAG-----TATTTGATGGCGCACTCCTTGTACGA 442
Db 291 TyrIleTyrrIleThrLeuGlnIlePheAspTrpMetAsnAlaValLeuThrLeuProGly 310
Qy 443 CTGCGTGGAGATTGTTCTGTATGAGGATGCGGATGCGATGCAATGCTTGTATGAA 502
Db 311 IleAlaIaLeuIleLeuGlyValGlyMetAlaValAspAlaAsnIleIleThrTyrrGlu 330
Qy 503 AGAATCCGAGAGAAATTTTATTTGTCTCAAAAGTCTTAAATAATGTGAGAAAAAGAT 562
Db 331 ArgIleYrGlnGluIleuLysSerValArgSerAlaPheArgSerGlyAsn 350
Qy 563 ACCAAGCTTTTGAGGCAATTTTGATTTCTAATCTTGAATAGTATGGCTCAGCACT 622
Db 351 ArgArgSerPheAlaThrIlePheAspAlaAsnIleThrIleIleAlaAlaVal 370
Qy 623 CTTTCTTCCAGATACAG36GCTTATTAAGGTTTGCTTGAATGATTTTGAAT 682
Db 371 LeuPheIlePheGlyThrSerSerValPheGlyPheAlaThrMetLeuIleLeuSerIle 390
Qy 683 TTCTCTCAATGTTTACGCTCTTTTGAATGATGCTTATTTTCTTATGCTG 733
Db 391 LeuThrSerPheIleThrAlaValPheLeuSerArgPheLeuAlaLeuLeuValGlu 410
Qy 734 -----TGATGAATAAG----- 745
Db 411 SerArgTrpLeuAspArgLysGlyTrpPheGlyValAsnLysLysIleMetAsp 430
Qy 746 -----ACCAACATACACAGTGTGATGATGATGATGATGATGATGATGATGAT 793
Db 431 IleGlnAspThrAspGluAsnThrGluProHisThrProPheGln-----Lys 446
Qy 794 CATGATTTCTTGAGAGATGCAAAAACTTGGCGCTGTTCTGAGAGTGTTTCTTTT 853
Db 447 TrpAspPheThrSerLysArgLysTyrrPhePheIlePheSerSerAlaValThrValAla 466
Qy 854 GGTTCGCTGCTCTCGGCTTTGAGCCTCGAATTCGTTTGGGAATGATTTAAAGA 913
Db 467 GlyIleIleIleLeuValPheArgLeuAsn-----LeuGlyIleAspPheAlaSer 484
Qy 914 GGG-----TATGCTTTTACCTTATTCACAAAAGACATGCGC 949
Db 485 GlyAlaIaArgIleGluValGlnSerAspHisLysLeuThrThrGluGlnValGluLysAsp 504
Qy 950 ATCAGCATGTTGCTCAAAATCGTGGCAAAAGTGTGCAATAAATCAAGAGAGCTGTCT 1009
Db 505 PheIleuSerLeuGlyMetAspProAspThrValIleLeuSerGlyGluLysSerAsnIle 524
Qy 1010 TCT--TCTAGAGACTTCGATTCAAAACATTTGATCTTCAGAAAAGATCAAAATCTAT 1066
Db 525 GlyValAlaIaArgPheValGlyValProAspLysGluThrIleAlaLysValLysThrTyrr 544
Qy 1067 TTTAGTATTAAGCTTTAAGCTATACTAAGACATACAGAGCTCTCTCTTAAATTAACG 1126

```

```

Db 545 PheLysAspLys----- 548
Qy 1127 ATCATGAGCTGGCGTTATTGTGGATTTGTTGTACAAAACGCGCTTGAATTTCTTACGA 1186
Db 549 ----- 550
Qy 1187 AACCTTAAACGAAACGCAAAATTTTGTCAAGGTAAAGACGAAACATATGAAAGAAAG 1246
Db 551 SerAspProAsnValSerThrValSerProThrValGlyLysGluLeuAlaArgAsnAla 570
Qy 1247 CGTTATCAGCGCAACATCGGCGCTTTTGAAGCTTTGGCAATCATCTTGTCTATGTAGT 1306
Db 571 LeuTyrrAlaValAlaIleAlaSerIleGly-----IleIleIleTyrrValSer 586
Qy 1307 TTGCGCTTGAATGCAAAATGCTTTCAAGTCCGTAATGCGCTTAAATTAATGACCTTTTG 1366
Db 587 IleArgPheGluTyrrYrMetAlaIleAlaIleAlaIleAlaSerLeuLeuTyrrAsp----- 604
Qy 1367 GCTACCGTGCAGCTCTTGTATTAAGCAATTTCTTTTGAAGAAATTCAAATAGATTGG 1426
Db 605 -----AlaPhePheIleValThrPhePheSerIleThrArgLeuGluValAspVal 621
Qy 1427 CAAGCATTTGCTTAAATGACTATTTGGGCTATTCATTAACATATCTTGTATCAT 1486
Db 622 ThrPheIleAlaIleAlaIleLeuThrIleIleGlyTyrrSerIleAsnAspThrIleValThr 641
Qy 1487 TTGATGCTATTTGCTGAA-----GATCCCAAGCGAACCCTGTTTACCCTATGCA 1537
Db 642 PheAspArgValArgLysIleHisMetLysArgLysProLysThrPheAlaAspLeuAsn 661
Qy 1538 GTTTTGTATTAATGATGCGCTTCAAAAGACGTTTACGCGGACGCTTATGCAACAGCTAC 1597
Db 662 HisIleValAsnLeuSerLeuGlnGlnThrPheThrArgSerIleAsnThrValLeuThr 681
Qy 1598 ACTCTATCAGATTTGTTTAAATGCTTTTGTATTAAGCGGCTCTCTGCTTAAATTTGCA 1657
Db 682 ValValIleValAlaValAlaThrLeuLeuIlePheGlyAlaSerSerIleThrAsnPheSer 701
Qy 1658 TTTATTAATGACATTAAGGATTTCTTGAAGATTTTGTCTTATTAATGACACACT 1717
Db 702 IleAlaLeuLeuValGlyLeuLeuThrGlyValTyrrSerSerLeuTyrrIleAlaAlaGln 721
Qy 1718 CTGTGTG 1726
Db 722 IleTrpLeu 724

RESULT 2
US-09-134-001C-3689
; Sequence 3689, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3689
; LENGTH: 766
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3689

Alignment Scores:
Pred. No.: 1,45e-65 Length: 766
Score: 647.50 Matches: 165
Percent Similarity: 50.95% Conservative: 129
Best Local Similarity: 28.60% Mismatches: 204

```

Qy	1049	GAAGAATCAAAATCTATTATTAGTCATAAAGCTTTAAAGCTATATCTACTACGACATACGAGCC	1108
Db	533	AsnLys-----AsnAlaThrValGlnPheLysLys	542
Qy	1109	TCCTCTCTAAATTAACGATCATGAGCTGGCGTTATTGTGGAGTTGTTGTGCAGAAACAGG	1168
Db	543	AspLeuSerArg-----	546
Qy	1169	CCTAGATTCTCTACGGAAACTCTAAACGAAACGCAAAATTTTGGTCAAAG-----	1219
Db	547	-----GluGluAspAsnLysLeuSerAlaLysValLysSerGluPheGlyAsp	562
Qy	1220	-----GTAAAGCAGCAAACTATCGAAGAAATCGCTTATCAGCGCAGCC	1261
Db	563	AsnProGlnIleAsnThrValSerProLeuIleGlyGlnGluLeuAlaLysAsnAlaVal	582
Qy	1262	ATCGGGCTTTTAGAGCTTTGGCAATCATCTTTGCTCTATGTGAGTTTGGCGCTTTTGAATGG	1321
Db	583	ThrAlaLeuIleLeuAlaSerIleGlyIleIleIleIleTyValSerLeuArgPheGluTrp	602
Qy	1322	CAATATGCTTTTACGTCCGTATCGCTTTAAATTATCATGACCTTTTGGCTACCTGTGCGATC	1381
Db	603	ArgMetGlyLeuSerValLeuAlaLeuLeuHisAspValPhe-----	617
Qy	1382	TTGTTTATACACATTTCTTTTGAAGAAATTCAAATAGATTGTGCAAGCCATTGGTGCT	1441
Db	618	IleIleIleAlaIlePheSerLeuPheArgLeuGluValAspLeuThrPheIleAlaAla	637
Qy	1442	TTAATGACTGTATTGGGGTATTTCATTAAACAATACTTTTGATCATTTTTGTATCGTATTCGT	1501
Db	638	ValLeuThrIleValGlyTyrSerIleAsnAspThrIleValThrPheAspArgValArg	657
Qy	1502	GAGAT-----CGCCACGCAACCTGTTTACC-----CCTATGCATGTTTATAGTTAAT	1549
Db	658	GluAsnLeuHisLysValLysValIleThrHisThrAspGlnIleAspAspIleValAsn	677
Qy	1550	GATGCGCTTCAAAAGAGCTTCAGCGCAGCGTAATGACACAGCTACACACTTATCATGTT	1609
Db	678	ArgSerIleArgIleGlnThrMetThrArgSerIleAsnThrValLeuThrValValValVal	697
Qy	1610	TTGTTAATGCTTTTGTATTAGCGCGCTCCTCTGCTCTTTAAATTTTGCATTTTATTATGACC	1669
Db	698	ValValAlaIleLeuIleLeuGlyAlaProThrIlePheAsnPheSerLeuAlaLeuLeu	717
Qy	1670	ATAGGAAATCTCTAGGAACCTTTATCGTCTCTTTATATATGACCACTCTCTG	1720
Db	718	IleGlyLeuLeuSerGlyValPheSerSerIlePheIleAlaValProLeu	734

```

; LENGTH: 323
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-462-844-4

```



## Alignment Scores:

Pred. No.: 4,15e-33 Length: 323  
 Score: 364.00 Matches: 102  
 Percent Similarity: 49.52% Conservative: 53  
 Best Local Similarity: 32.59% Mismatches: 134  
 Query Match: 10.99% Indels: 24  
 DB: 4 Gaps: 9

US-09-868-987-1 (1-1864) x US-09-462-844-4 (1-323)

QY 794 CATGATTTCTGAGAGATGCAAAAACCTTGGCGTGTTCCTGGAAGTCTTTCTTTA 853  
 Db 17 TyrAspPheMetArgTPrAspTyrTPrAlaPheGlyLeuSerGlyLeuLeuLeuIleAla 36  
 QY 854 GGTTCGCTGCTCGGGTGTGAGCTGGAATTCCTTTGGAAATGAGATTTTAAAGGA 913  
 Db 37 AlaIleValIleMetGlyValArgGlyPheAsn-----TPrGlyLeuAspPheThrGly 54  
 QY 914 GGGTAAGCTTACCTTAAATCCAAAGAGATGGACACAGCATGTTGCTCAAAATGGCT 973  
 Db 55 GlyThrValIleGluIleThrLeuGlyLeuProAlaGluIleAspVal-----MetArg 72  
 QY 974 GGCAAGTTGTGCAATTAACCTACAGAAAGCTGCTTCTTCTAGAGACTTCCGATTCAA 1033  
 Db 73 -----AspAlaLeuGlnIleValAspIlePhe-----GluGluProMetLeuGln 86  
 QY 1034 ACATTTGGATCTTCAGAAAAGATCAAAATCTATTAGTGAATAAGCTTAAGCTTAAT 1093  
 Db 87 AsnIleGlySerSerHisIleAspIleMetValArgMetProProAlaGluGlyIleThrGly 106  
 QY 1094 AAGCAGATACAGACCTCT---CTCCTAAATTAACGATACAGATGCGCGCTTATTGGGG 1150  
 Db 107 GlyIleValIleGlySerGlnValLeuValIleAsnGluSer----- 121  
 QY 1151 ATGTGTTCAGAAACAGCGCTAGATTTCTACGGAACCTTAAACGAAACGAAATTT 1210  
 Db 122 -----ThrAsnGlnAsnAlaIleValIleValArg 130  
 QY 1211 TGGTCAAGGTAAGACAGAACTATCGAAGAAATCGCTTATCAGCGACATCGGGCTT 1270  
 Db 131 IleGluPheValGlyProSerValGlyValAspLeuAlaGlnThrGlyAlaMetAlaLeu 150  
 QY 1271 TTAGAGCTTTGGCAATATCTGCTCTATGTGAGCTTGGCTTGGATGGCAATATGCT 1330  
 Db 151 MetAlaIleLeuLeuSerIleLeuValIleValGlyPheArgPheGluIleTPrArgLeuAla 170  
 QY 1331 TTCAGTCCGATGCGCTTATATTCATGACCTTTGGCCTGTCGACGCTTGTATTATA 1390  
 Db 171 AlaGlyValValIleAlaLeuAlaHisAspValIleIleThrLeuGlyIleLeuSerIleu 190  
 QY 1391 GCACATTTCTTTTGAAGAAATTCAAATAGATTGGCAAGCATTTGGCTTTATGACT 1450  
 Db 191 PheHis-----IleGluIleAspLeuThrIleValAlaSerIleuMetSer 205  
 QY 1451 GTATTGGGGTATCTTAACAATATCTTGATCATTTTGTATCGATTCGGAGAT--- 1507  
 Db 206 ValIleGlyTyrSerLeuAsnAspSerIleValAlaSerAspArgIleArgGluAsnIle 225  
 QY 1508 CGCAAGGAAACCTGTTAACCCCTATGATGATGTTAGTTAATGATGACCTTCAAAAGAG 1567  
 Db 226 ArgIleValIleArgArgGlyThrProTyrGluIlePhe---AsnValSerLeuThrGlnThr 244  
 QY 1568 TTCAGCCGACGTAATGACAACAACCTCAACTATACAGTTTGTATTGTTGTTT 1627  
 Db 245 LeuHisArgThrIleuIleThrSerGlyThrThrLeuMetValIleLeuMetLeuTyrIleu 264  
 QY 1628 ATAGCGGCTCTCTGCTCTTAATTTTGATTTATATGACATATGAGATCTTCTAGGA 1687  
 Db 265 PheGlyGlyProValIleuGluGlyPheSerLeuThrMetLeuIleGlyValSerIleGly 284  
 QY 1688 ACTTATGCTCTCTTATATGACACACCTCGTGTGTG 1726  
 Db 285 ThrAlaSerSerIleTyrValAlaSerAlaLeuAlaLeu 297

## RESULT 4

US-09-462-844-3

/ Sequence 3, Application US/09462844  
 / Patent No. 6258563  
 / GENERAL INFORMATION:  
 / APPLICANT: Quax, Wilhelmus J.  
 / TITLE OF INVENTION: Increasing Production of Proteins in Gram-Positive Microorganisms  
 / FILE REFERENCE: GC385-US  
 / CURRENT APPLICATION NUMBER: US/09/462,844  
 / CURRENT FILING DATE: 2000-01-13  
 / PRIOR APPLICATION NUMBER: PCT/US98/14786  
 / PRIOR FILING DATE: 1998-07-15  
 / PRIOR APPLICATION NUMBER: EP 97305286.3  
 / PRIOR FILING DATE: 1997-07-16  
 / PRIOR APPLICATION NUMBER: EP 97305344.0  
 / PRIOR FILING DATE: 1997-07-17  
 / NUMBER OF SEQ ID NOS: 11  
 / SOFTWARE: FastSeq for Windows Version 4.0  
 / SEQ ID NO 3  
 / LENGTH: 615  
 / TYPE: PR1  
 / ORGANISM: Escherichia coli  
 / US-09-462-844-3

## Alignment Scores:

Pred. No.: 1.92e-32 Length: 615  
 Score: 359.50 Matches: 80  
 Percent Similarity: 62.39% Conservative: 61  
 Best Local Similarity: 35.40% Mismatches: 82  
 Query Match: 10.85% Indels: 3  
 DB: 4 Gaps: 2

US-09-868-987-1 (1-1864) x US-09-462-844-3 (1-615)

QY 44 ATGACACATATCTGCAACCGTGGATGCGTATGAGTGAATGACGGTATATG 103  
 Db 375 SerGlyValIleValAspAlaAsn-----GlyArgAlaValLeuValIleGlnGluVal 392  
 QY 104 GTCGAGCCCTATTTTAAACGTCCATTTGAAATTCATGCCAGTCTTCAGGAAATTT 163  
 Db 393 IleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleAsn 412  
 QY 164 ACCCAGCGTAAGTGAACCAACTCGCTCGATTTAAATCTGAGCGATGCTTTGTT 223  
 Db 413 AspProAsnGluIleAlaArgGlnLeuSerLeuLeuLeuValArgIleValIleAlaPro 432  
 QY 224 CCGAGGTTCTGACGTAGAGACAGATCTCTGATCTTGGAAAAACAATGATACAA 283  
 Db 433 IleGlnIleValIleGluIleGluArgThrIleGlyProThrLeuGlyMetGlnAsnIleGluGln 452  
 QY 284 GGCATTTCTCAGCATGCTGCTGGCAATGCTTATTTGATGGACGATATATAT 343  
 Db 453 GlyLeuGluIleValIleValIleGluIleGluValSerIleLeuPheMetIleIlePheTyr 472  
 QY 344 AGATTGGAGCGCTATCGCTTGGAGAGCTTCTTGTGATCTTTGCTTATCTGGGCA 403  
 Db 473 LysIlePheGlyLeuIleAlaThrSerAlaLeuIleAlaAsnLeuIleuIleValGly 492  
 QY 404 GCTGTACAGTATTG---GATGGCCCATCGACCTTGCAGACCTCGGTGGATTGTTCTT 460  
 Db 493 IleMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512  
 QY 461 GCTATGGAGTGGCCGTAGATGCAAAATGTTCTGTATTCGAAGATCCAGAGGAATTT 520  
 Db 513 ThrIleuAlaValAlaValAspAlaAsnValLeuIleAsnGluArgIleIleValGluLeu 532  
 QY 521 TTATTGCTCAAAAGTCTTAAATAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCC 580  
 Db 533 SerAsnGlyArgThrValIleGlnIleAlaIleAspGluGlyTyrArgGlyAlaPheSerSer 552  
 QY 581 ATTTTATTTCTAATCTGATCAGTATTTGGCTTCAGACACTTCTTTCTTCATGATACA 640





```

QY 1013 TCTAGAGACTTCGGTATT-----CAAACTTTGGATTTCAGAAAGATCAA 1060
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 SerileaspVallysalVallysglnasnlsglnThrpsrSerSerGlnlylelys 796
QY 1061 ATCTATTATTGATTAAGCTTTTAAGCTATTAAGCAATAGCA-----GCCCTCTC 1114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 797 AsnilleusleusnlyspProthrGlyGlyThrilleusleusSerGluilleaSerleu 816
QY 1115 -----CTAAATTACGATCATGAGCTGGCGCTTAATTGGGATTGTT 1156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 817 LyselyserThrThrproserlyleuThrGlnGlnaspGlyAspIlyrAlaThrThrVal 836
QY 1157 GTC---AGAAACAGCGCTAATTTCTCTACGAAACTCTAAACGAAACGAAATTGG 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 837 ThrGlylyserValThrAspIlyAspValGlyGlylyserSerGln-----GlnValMet 853
QY 1214 TCMAAGGTAAAGCAACATTCGAAAGAAATGCGTTATCAGCGCAACATCGG----- 1267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 854 AlalysVallyAsp---LeuGlnlyspProserHisileysIleasnValGlyGlyAla 872
QY 1268 -----CTTTTAGAGCTTTTGCA 1285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 873 ThrAspAspIleAspIlysalalleSerGlnleuAlaMetIleAlaIleIle 892
QY 1286 ATCATCTTGCTCTATGATGATTTGCGCTTGGAATGCAATATGCT---TTCAGTGGCGTA 1342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 893 IleValIyIreulleuValIleThrPheArgIlyGlyleuAlaProPheThrIleIeu 912
QY 1343 TGGCCTTAATTCATGACCTTTGGGCAACCGTGACAGCTGTTGTTATAGCAATTTCTTT 1402
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 913 PheSerleuProPheThrValIleGlyValValleuAlaMetIleIleThrGly----- 930
QY 1403 TTGAAGAAATTCATAGATTTGCAAGCCATGTGCTTTAATGACTGTATTTGGGGTAT 1462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 931 ---GluThrIleSerValPro---SerleuIleGlyMetleuMet---LeuIleGlyIle 947
QY 1463 TCATTAAACAATCTTGATCTTTGATGATGATGCTGATTCGTGAAGATCCGCAACGACCTG 1522
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 948 ValValIleThrAsnAlaIleValIleuIleAspArgValIleAsnAsnGlnIlysglnIlyMet 967
QY 1523 TTTACCCCTATGCATGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1582
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 968 -----ProMetIlysglnAlaIleuIleGlnIleGlyIleThrArgIleAspProIleIeu 985
QY 1583 ATGACACAGCTACCACTCATCTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 986 MetThrAlaIleAlaThrIleGlyAlaIleuValProleuIlePheGlnHisAspSerSer 1005
QY 1643 GTCCTT-----AATTTTCATTTATATGACCATGAGGATTTCTTAAGAACTTTA 1693
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1006 IleleuIleSerIlysglyMetAlaIleThrValIleGlyIleuIleSerSerThrIleu 1025
QY 1694 TCGTCTCTTATATGACCACTCTG-----TTGTTGTTTATGATGATGATGATGATGAT 1747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1026 LeuThrIleuPheValIleProValIleIleGlyIleuIlePheThrIleuIleAsnIlyle 1045
QY 1748 CGCTCAAA 1756
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1046 ThrIlylys 1048

```

```

RESULT 6
US-09-022-699-2
; Sequence 2, Application US/09022699
; Patent No. 6060277
; GENERAL INFORMATION:
; APPLICANT: KIKLY, KRISTINE
; APPLICANT: SOUTHAN, CHRISTOPHER
; APPLICANT: KNAB, ANNE
; TITLE OF INVENTION: Human AFCL
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA

```

```

; STREET: P. O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/022,699
; FILING DATE: 12-FEB-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 97304440.7
; FILING DATE: 12-JUN-97
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-022-699-2

Alignment Scores:
Pred. No.: 0.00043 Length: 474
Score: 117.00 Matches: 55
Percent Similarity: 41.55% Conservative: 31
Best Local Similarity: 26.57% Mismatches: 67
Query Match: 3.53% Indels: 54
DB: Gaps: 9

US-09-868-987-1 (1-1864) x US-09-022-699-2 (1-474)
QY 1178 CTCACGAAATCTTAACGAAACGAAATTTGGTCAAGGTAAAGCAACATCATG 1237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69 LeuIyrcIleuAspIlySerThrPheSerPheThrIleu----- 83
QY 1238 AAGAAATGCGTTATGACGCGCATCGGGCTTTTGAAGCTTTGGCAATCATCTGCTC 1297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 -----TyrSerGlnThrGlnGlyThrIleu-----IleleuIeu 94
QY 1298 TATGTAGTTGGCGCTTGAATGCA-----TATGCT----- 1330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 PheGlyGlyIleProIyIreuThrArgIleuSerGlyArgPheCysGlyIyrAlaGlyPhe 114
QY 1331 -----TTCAGTGGCGGTAGCGCTTATTCATGACCTTTGGCTACCTTGACATC 1381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 GlyProGlnIyrcIleuIleThrGlnSerIleuValPheleuIleuAlaThr----- 131
QY 1382 TTGTTTATGACATTTCTTTTGAAGAAATTCMAATAGATTGCAAGCATTTGGTCT 1441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 LeuPheSerAla----- 135
QY 1442 TTAATGCTGATTTGGGATTTCAATTAACAATCTTTGATCGTATTCGT 1501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 ---LeuThrGlyIleuProIyPserIleuIyrAsnThrPheValIleGlnGlnIlyshIsgly 154
QY 1502 GAAGATGCCAAGCAACCTGTTTACCCCTTATGATGATGATGATGATGATGATGATGATGATGAT 1561
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 PheAsnGlnGlnThr-----LeuGlyPhePheMetArgAspAlaIleIlys 169
QY 1562 AAGACGTTAGCGCGCAAGGTATGACAAACGCTACCACTATCAAGTTTGTAAAGCTT 1621

```

Db 170 Lys---PheValValThrGlnCysIleLeuLeuProValSerSerLeuLeuLeuTyrIle 188  
 Qy 1622 TTGTTTATAGCGGCTCTCTGCTTTTAAATTTTGCATTATATGACC-----ATAGGG 1675  
 Db 189 IleLysIleGlyGlyAspTyrPhePheIleTyrAlaTrrPhePheThrLeuValSer 208  
 Qy 1676 ATTCTTCTAGAACCTTATCGTCTCTTATATATGACACCTCTGTTGTTATGCTC 1735  
 Db 209 LeuValLeuValThrIleTyrAlaAspTyrIleAlaProLeuPheAspLysPheThrPro 228  
 Qy 1736 CGTAAAGAAATCGCTCAAAA 1756  
 Db 229 LeuProGluGlyLysLeuLys 235

## RESULT 7

US-09-134-001C-5389  
 ; Sequence 5389, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 5389  
 ; LENGTH: 901  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-5389

## Alignment Scores:

Pred. No.: 0.00219 Length: 901  
 Score: 108.00 Matches: 106  
 Percent Similarity: 36.58% Conservative: 108  
 Best Local Similarity: 18.12% Mismatches: 227  
 Query Match: 3.26% Indels: 144  
 DB: 4 Gaps: 28

US-09-868-987-1 (1-1864) x US-09-134-001C-5389 (1-901)

Qy 137 TTAATCTGGAGCATGCTCTTTGTTCCGAGGTCTCAGTGAAGACGATCTCTCT 256  
 Db 147 LeuLysAspSerSerLysHisIleIleAspLysGluLeuLysAspValThrAspAsnHis 166  
 Qy 257 GATCTT-----GGGAAAAACAATGTACACAAGGCATTATCTCA 295  
 Db 167 AsnValGlnIleGluLysThrGlnGlyAlaMetAsnSerGluProGlyGlyThrSer 186  
 Qy 286 GCATGCTGTGGCTTGGCAATGCTTATGTTTGCAGCGTATATATAGATTTCGAGGC 355  
 Db 187 GluIleValGlyIleIleValAlaPheValIleLeuLeuLeu-----ThrPheGlySer 204  
 Qy 356 GTCATCGCTTCGGAGCTGTTCTTCTGATCTTTTGCTT----- 394  
 Db 205 LeuIleAlaAlaGlyMetProIleIleSerAlaIleIleGlyLeuGlySerValGly 224  
 Qy 395 ATCTGGGAGCTCTACAGTATTG---GATGGCCA-----CTCACCTTGTGAGGA 442  
 Db 225 IleIleAlaLeuLeuThrTyrIlePheAspIleProAsnPheThrLeuThrLeuAla--- 243  
 Qy 443 CTCGCTGGGATGTTCTGCTATGGGATGCCCGTAGATGCAAAATGTTCTTGTATCGAA 502  
 Db 244 -----ValMetIleGlyLeuAlaValGlyIleAspTyrSerLeuPheIleLeuPhe 260  
 Qy 503 AGAATCCGAGCATTTTATTGCTCAAGCTTAAAAA----- 544  
 Db 170 Lys---PheValValThrGlnCysIleLeuLeuProValSerSerLeuLeuLeuTyrIle 188  
 Qy 1622 TTGTTTATAGCGGCTCTCTGCTTTTAAATTTTGCATTATATGACC-----ATAGGG 1675  
 Db 189 IleLysIleGlyGlyAspTyrPhePheIleTyrAlaTrrPhePheThrLeuValSer 208  
 Qy 1676 ATTCTTCTAGAACCTTATCGTCTCTTATATATGACACCTCTGTTGTTATGCTC 1735  
 Db 209 LeuValLeuValThrIleTyrAlaAspTyrIleAlaProLeuPheAspLysPheThrPro 228  
 Qy 1736 CGTAAAGAAATCGCTCAAAA 1756  
 Db 229 LeuProGluGlyLysLeuLys 235  
 Db 261 ArgPheLysGlu-----LeuLysLysLysGlyValAspThrVal 273  
 Qy 545 ---TCTGTAGAAAAAGGATATACCAAGCGCTTTTGGAGCCATTTTTGTATTCTAACTTGACT 601  
 Db 274 GluAlaIleAlaThrAlaValGlyThrAlaGlySerAlaValIlePheAlaGlyLeuThr 293  
 Qy 602 ACAGTATTGGCC-----TCAGCACTTCTTTCTTCAGTACACAGGCGCTATTAAAGGG 655  
 Db 294 ValMetIleAlaValCysGlyLeuSerLeuValGlyIleAspPheLeuAlaValMetGly 313  
 Qy 656 TTT-----GCTTTGACATTGATTATTAGGAATTTTCTCTCTCAATG----- 694  
 Db 314 PheAlaSerAlaIleSerValLeuPheAlaValLeuAlaIleLeuThrLeuLeuProAla 333  
 Qy 695 TTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAATAAGACCCCAACAT 754  
 Db 334 LeuIleSerIlePheHisLysSerIleLysIleLysAspLysProThrLysSerLysAsp 353  
 Qy 755 ACACAGTTGCATATGATGATTAAGTTC---GTGGGATAAAGCATGATTTCTTGAGAGGA 811  
 Db 354 ProLysAspHisSerTrpAlaLysPheIleValGly----- 365  
 Qy 812 TGCAAAAAATCTTGGCGTGTCTGGAAGTCTTTTCTTTTAGCTTGGTGTCTCTC--- 868  
 Db 366 ---LysProValIleAlaValIleValSerLeuIleLeuLeuLeuAlaIlePro 384  
 Qy 869 -----GGG 871  
 Db 385 ValSerGlyMetArgLeuGlyIleProAspAspSerLeuLysProThrAspSerSerGlu 404  
 Qy 872 TTTGGAGCTTGGAAATTCGGTT-----TTGGGAATGGATTTTAAAGGAGGTATGCC 922  
 Db 405 TyrLysAlaTyrLysLeuIleSerAspAsnPheGlyGlyTyrAsnGlyGlnIleVal 424  
 Qy 923 TTTACCTTTAATCCAAAGAGCATGGCATCAGC-----GATGTTGCTCAA 967  
 Db 425 MetLeuValAsnThrLysAspGlyGlySerLysSerThrIleGluArgAspLeuAsnAsn 444  
 Qy 968 ATGCGTGGCAAGTT-----GTGCATAAATACAGGAAGCTGTGCTTTCT 1012  
 Db 445 MetArgSerAspLeuGluAspIleAspAsnValAspThrValSerLysAlaGlnLeuThr 464  
 Qy 1013 ---TCTAGAGACTTCCGTATTCAAAACATTT-----GGATCT 1045  
 Db 465 AspAsnAsnAsnTyrAlaLeuPheThrIleIleProGluLysGlyProAsnSerGlnSer 484  
 Qy 1046 TCAGAAAGATCAAAATCTATTTTAGTGATAAAGCTTTAAGCTATATCTAAGCAGATACGA 1105  
 Db 485 ThrGluAsnLeuValTyrAspLeuArgAsp-----TyrHisSerGlnAla--- 499  
 Qy 1106 GCCTCTCTCTCTAAATTAACGATCATGAGCTGGCGTTATTGTGGGATTGTTGTCAGAAAC 1165  
 Db 499 ----- 499  
 Qy 1166 AGGCTAGATTTCTCTACGAAACTCTAAACGAACGCAAAATTTTGGTCAAGGTAGC 1225  
 Db 500 GlnGluLysTyrAspTyrGlyThrGluIleSerGlyGln-----SerValIleAsn 516  
 Qy 1226 AGCAAACTATCGAAGAAATGCGTTATCAGCGGACCATC-----GGGCTTTAGGAGCT 1279  
 Db 517 IleAspMetSerGluLysLeuAsnAsnAlaIleProValPheAlaGlyValIleVal 536  
 Qy 1280 TTGGCAATCATCTTGTCTATGTGAGTTTGGCTTTTGAATGGCAATATGCTTTTCAGTGCC 1339  
 Db 537 LeuAlaPhePheLeuLeuMetIleValPheArg---SerIleLeuValProLeuLysAla 555  
 Qy 1340 GTATGCGCTTTAAATCATGACCTTTTGGCTACCTGT-----GCAGTCTCTTTATAGCA 1393  
 Db 556 ValLeuGlyPheIleLeuSerLeuMetAlaThrLeuGlyPheThrLeuValIleGln 575  
 Qy 1394 CATTTCTTTTGAAGAAAAATTCAAATAGATTGCAAGCATTTGGTCTTTTATGACTGTA 1453  
 Db 576 HisGlyPheMetGlySerLeu---PheGlyIleGluAsnThrGlyProLeuLeuAlaPhe 594



```

QY 1355 CATGACCTTTGGCTACGCTGCTGCTTATAGACATTTCTTTTGAAGAAATT 1414
Db 282 GlyCysLysAlaLysThrCysAspIleuLeuThrLysAlaGluTyrPheLeu 298
QY 1415 CAAATAGATTTCGAAGCCATTGGTGGCTTTAACTGACTGATTGGGGTATTCAATTAAACAAT 1474
Db 299 -----ValLeuAlaValLeuAsnSerGlyThrAsnPro 309
QY 1475 ACTTTGATCATTTTTCATCGTATTCTGTGAAGATCGCAAGCAACCTGTTTACCCTATG 1534
Db 310 IleIle----- 311
QY 1535 CATGTTTGTAGTTAATGATGCCCTTCAAAGACGTTTCAGCCGACCGTA 1582
Db 312 TyrThrLeuThrAsnLysGluMetArgAlaPheIleArgIle 327

RESULT 9
US-08-760-936-4
; Sequence 4, Application US/08760936
; Patent No. 5856443
; GENERAL INFORMATION:
; APPLICANT: MacLennan, A. John
; TITLE OF INVENTION: Molecular Cloning and Expression of
; TITLE OF INVENTION: G-Protein Coupled Receptors
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/760,936
; FILING DATE: December 6, 1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Doran R.
; REGISTRATION NUMBER: 38,261
; REFERENCE/DOCKET NUMBER: MAC-100C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 383 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-760-936-4

Alignment Scores:
Pred. No.: 0.00459 Length: 383
Score: 103.50 Matches: 86
Percent Similarity: 31.80% Conservative: 59
Best Local Similarity: 18.86% Mismatches: 132
Query Match: 3.12% Indels: 179
DB: 2 Gaps: 22

US-09-868-987-1 (1-1864) x US-08-760-936-4 (1-383)
QY 248 ATCTCTTCGTGCGGAAACAAATGTACAAAGCATTTATCTACGATGCTGTGCG 307
Db 40 ValGluLysAspHisGlyIleLysLeuThrSerValValPheIleLeuLysCys 58
QY 308 TTGGCAATGCTTATGTTTGTATGACGCTATATATAGATTGGAGGCGTCATCGCTCG 367

```

```

Db 59 -----LeuIleIleLeuGluAsnIlePheVal----- 67
QY 368 GGAGCTGTTCTTCAATCTTTTGTCTTATCTGGGAGCT-----CTA 409
Db 68 -----LeuLeuThrIleTrpLysThrLysLysPheHisArgProMet 81
QY 410 CAGTATTTGGATCGCCACTCCTACCTTGTGACGACTCGCTGGGATTTCTTGTCTATGGGG 469
Db 82 TyrTyrPheIleGlyAsnLeuAlaLeuSerAspLeuLeuAla-----Gly 96
QY 470 ATGGCCGTAGATGCAAAATGTTCTTATTCGAAAGAAATCCGAGAGGAATTTTATTGTCT 529
Db 97 ValAlaTyrThrAlaAsnLeuLeu----- 105
QY 530 CAAAGTCTTAAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTAT 589
Db 105 ----- 105
QY 590 TCTAACTTGACTACAGTATTGGCCTCAGCAGCTTCTTTTCTCTAGATACAGGCGCTATT 649
Db 106 SerGlyAlaIleThrTyrLysLeuThrProAlaGlnTrpPheLeuArgGluGly----- 123
QY 650 AAAGGGTTTGTCTTGGACATTCATTTTAGGAATTTTCTCTCAATGTTTACG-----GCT 703
Db 124 -----SerMetPheValAlaLeuSerAlaSerValPheSerLeuLeuAla 138
QY 704 CTTTTCATGACTAAATTTTCTTCTCATGCTGTGGATGAATAAGACCCCAACATACACAGTTG 763
Db 139 IleAlaIleGluArgTyrIleThrMetLeuLysMetLys-----LeuHisAsnGlySer 156
QY 764 CATATGATGAATAAGTTCGTGGGATGAAGCATGATTTCTTGAGAGGATGCAAAAACCTT 823
Db 157 AsnSerSerArgSerPheLeu-----LeuIleSerAlaCys----- 168
QY 824 TGGGCTGTTCTGGAAGTGTTTCTTTTAGTTCGCTGCTCGGTTTGGAGCCCTGG 883
Db 169 TrpValIleSer-----LeuIleLeuGly-----GlyLeuProIleMetGlyTrp 183
QY 884 AATTCGTTTGGGAATGGATTTTAAAGAGGGGTATGCTTTTACCTTTTAAATCCAAAGAG 943
Db 184 AsnCysIle----- 186
QY 944 CATGGCATCAGCGATGTTGTCATAATGCGTGGCAAAAGTTGTGCATAAACTACAGGAAGCT 1003
Db 187 -----Ser 187
QY 1004 GGTCTTCTCTAGAGACTTCCGTATTCAAACATTTGGATCTTCAGAAAAGATCAAAATC 1063
Db 188 SerLeuSerSer-----CysSerThrValLeuProLeu 198
QY 1064 TATTTTAGTGATAAGCTTTAAGCTATATAAGCAGATACGAGCTCTCTCTCTAAATTA 1123
Db 199 TyrHisLysHisTyrIleLeuPheCysThrThrValPheThrLeuLeuLeuSerIle 218
QY 1124 ACATATCATGAGCTGCGGTATTGTGGGATTT-----GTTGTCAAGAACAGCCCTAGATTT 1177
Db 219 ValIleLeu-----TyrCysArgIleTyrSerLeuValAlaGlnThrArgSerArgArg 235
QY 1178 CTCTACGGAAACTCTAAACGNAACGCAAAATTTTGGTCAAGAGTAAGCAGCAAACTATCG 1237
Db 236 Leu-----ThrPheArgLysAsnIleSerLysAlaSerArgSerGluLys----- 251
QY 1238 AAGAAATATGCGTTATCAGCGCAGCATCGGGCTTTTAGAGGCTTTTGGCAATCATCTTG--- 1294
Db 252 -----SerLeuAlaLeuLeuLysThrValIleIleValLeuSer 264
QY 1295 CTCTATGATGAGTTGCGCTTTGAATGGCAATATGCTTTCAGTGCCTGCTGCTTTAATT 1354
Db 265 ValPheIleAlaCys-----TipAlaProLeuPheIleLeuLeuLeuAspVal 281
QY 1355 CATGACCTTTTGGCTACGCTGCTGCTGCTTTGTTATAGACATTTCTTTTGAAGAAATT 1414
Db 282 GlyCysLysAlaLysThrCysAspIleLeuThrLysAlaGluTyrPheLeu----- 298

```

```

OY 1415 CAATAGATTGGCAAGCATTGGCTTATGACTATTTGGGGATTTCATTAACAAT 1474
DB 299 -----ValleuAlaValleuAsnSerGlyThrAspPro 309
OY 1475 ACTTGATCATTTTGTATGATTCGTGAAGATCGCAAGCAACCTGTTACCCCTATG 1534
DB 310 IleIle----- 311
OY 1535 CATGTTTACTATATGATGCGCTTCAAAAGACGTTGACCGCACGGTA 1582
DB 312 TyThrIleuThrAsnLysGluMetArgArgAlaPheIleArgIleIle 327

RESULT 10
US-08-845-566-3
; Sequence 3, Application US/08845566
; Patent No. 5912144
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Guegler, Karl
; TITLE OF INVENTION: EDG-1 LIKE RECEPTOR
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,566
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PP-0271 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SRO ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 181948
; US-08-845-566-3

Alignment Scores:
Pred. No.: 0.00888 Length: 381
Score: 101.00 Matches: 83
Percent Similarity: 31.83% Conservative: 58
Best Local Similarity: 18.74% Mismatches: 116
Query Match: 3.05% Indels: 186
Gaps: 22

US-09-868-987-1 (1-1864) x US-08-845-566-3 (1-381)
OY 290 ATCTCAGATGCTGTGGCTGGCAATGCTTATGTTGGATGAGCGATATTAATGATT 349
DB 53 IleIleuIleCysCys-----PheIleIleLeuGluAsnIlePheVal----- 66
OY 350 GGAGGCGTCATCGCTTGGGAGCTGTCTTCGAAATCTTTGCTATATCTGGGCACT--- 406

```

```

DB 67 -----LeuLeuThrIleTrpLysThrLys 74
OY 407 -----CTACAGATTTTGGATGGCCACATCAGCTTCTGACGAGCTTCGCTGG 451
DB 75 LysPheHisArgProMetLysTrpPheIleGlyAsnLeuAlaLeuSerAspLeuAla 94
OY 452 ATGTGTTCTGATGAGGATGAGCGGTAGATGCAAAATGTTCTGTATGCAAAATCCGA 511
DB 95 -----GlyValAlaIleThrAlaAsnLeuLeu----- 104
OY 512 GGAGATTTTATTTGTCGCAAGCTTTAAAAATCTGTAGAAAAAGATATACCAAGCT 571
DB 104 ----- 104
OY 572 TTGGAGCCATTTTGATTTCTAATCTGACTACAGATTGGCCTCAGCACTTCTTTCTTC 631
DB 105 -----SerGlyAlaThrThrThrLysLeuThrProAlaGlnTrpPhe 118
OY 632 CTAGATACAGGCGCTATTAAGGTTTGCTTGACATGATGATTAGCAATTTTCTCTCA 691
DB 119 LeuArgGluGly-----SerMetPheValAlaLeuSerAlaSer 131
OY 692 ATGTTTACG-----GCTCTTTGATGACTAAATTTTCTCATGCTGTGATG----- 739
DB 132 ValPheSerIleuLeuAlaIleAlaIleGluArgGlyIleThrMetLeuLysMetLysLeu 151
OY 740 ---AATNAGACCCACATACAGATGATGATGATGATGATGATGATGATGATGATGAT 796
DB 152 HisAsnGlySerAsnAsnPheArgLeuPheLeuLeu----- 163
OY 797 GATTTCCTGAGAGATGCAAAAACTTGGGCTGTTCTGGAAGTGTTCCTTTTATAGT 856
DB 164 -----IleSerAlaCys-----TrpValIleSer-----LeuIleGluGly 175
OY 857 TGGGTTGCTCTCGGTTTGGAGCTGGAATCCGTTTGGGAATGATTTAAAGAGG 916
DB 176 -----GlyLeuPProIleMetGlyTrpAsnGlyIle----- 185
OY 917 TATGCCCTTACCTTAAATCCAAAGAGCATGAGCATGAGCATGTTGCTCAATGCGTGC 976
DB 185 ----- 185
OY 977 AAGTTGTCATTAACATACAGAGAGCTGCTTCTTCTAGAGACTTCGATTAACAACA 1036
DB 186 -----SerAlaLeuSerSer----- 190
OY 1037 TTGATCTTCAGAAAAAGATCAAAATCTATTTGATGATTAAGCTTAAAGCTTAAGT 1096
DB 191 -----CysSerThrValLeuProLeuTyrHisIleLysIleTyrIleLeuPheCysThrThr 208
OY 1097 CAGATACAGACCTCTCTCCATAAATTACATGATGAGCTGGCTTATCTGGATT--- 1153
DB 209 ValPheThrLeuLeuLeuLeuSerIleValIleLeu-----TyrCysArgIleTyr 225
OY 1154 ---GTTGTCAGAAACAGGCTAGATTCTCTACGAAACTCTAAGCAAAAGCAAAATTT 1210
DB 226 SerIleuValArgThrArgSerArgArgLeu-----ThrPheArgLysAsn----- 240
OY 1211 TGGTCAAAAGGTAAAGCAAACTATGAGAAAAATGCGTTATGAGCGACATCGGCGTT 1270
DB 241 -----IleSerLysAlaSerArgSerArgSerGluAsnValAlaLeu 253
OY 1271 TTAGAGCTTTGGCAATCATCTTG---CTCATGATGAGTTTGGGCTTGAATGCAATAT 1327
DB 254 LeuLysThrValIleIleValLeuSerValPheIleAlaCys-----TrpAlaPro 270
OY 1328 GCTTACAGTGGCGATGCGCTTAAATCATGACCTTTGGGCTACCTGAGCGATGCTGTT 1387
DB 271 LeuPheIleLeuLeuLeuLeuAspValGlyCysLysValLysThrCysAspIleLeuPhe 290
OY 1388 ATACACATTTCTTTTGAAGAAAAATTCMAATGATTTGCAAGCATTTGGTCTTATATG 1447
DB ----- 1447

```

Db 291 ArgAlaGluTyrPheLeu-----Valleu 298  
Qy 1448 ACTGATTTGGGTATTCATTAACAATACTTTGATCACTTTTGTGATCGTATTCGTAAGAT 1507  
Db 299 AlaValLeuAsnSerGlyThrAsnProIlelle----- 309  
Qy 1508 CGCCAAGCGAACCTGTTTACCCTCATGATGTTTGTAGTTAATGATGCCCTTCAAAAGACG 1567  
Db 310 -----TyrThrLeuThrAsnLysGluMetArgAla 320  
Qy 1568 TTCAGCCGC 1576  
Db 321 PheIleArg 323  
RESULT 11  
US-08-467-948A-28  
; Sequence 28, Application US/08467948A  
; Patent No. 5998164  
; GENERAL INFORMATION:  
; APPLICANT: LI, YI  
; APPLICANT: CAO, LIANG  
; APPLICANT: NI, JIAN  
; APPLICANT: GENTZ, REINER  
; APPLICANT: BULT, CAROL J.  
; APPLICANT: SUTTON III, GRANGER G.  
; APPLICANT: ROSEN, CRAIG A.  
; TITLE OF INVENTION: Polynucleotides Encoding Human G-Protein  
; TITLE OF INVENTION: Coupled Receptor GPR2  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
; STREET: 1100 NEW YORK AVE., NW, SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PATENTIN RELEASE #1.0, VERSION #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,948A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/04079  
; FILING DATE: 30-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEFFE, ERIC K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488.1140003/EKS/KLM  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 381 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
US-08-467-948A-28  
Alignment Scores:  
Pred No.: 0.00888 Length: 381  
Score: 101.00 Matches: 83  
Percent Similarity: 31.83% Conservative: 58  
Best Local Similarity: 18.74% Mismatches: 116  
Query Match: 3.05% Indels: 186  
DB: 2 Gaps: 22  
US-09-868-987-1 (1-1864) x US-08-467-948A-28 (1-381)

Qy 290 ATCTCAGCATGCTGTGGCTTGGCAATGCTTATTGTTTGTAGAGCGTATATATAGATT 349  
Db 53 IleLeuIleCysCys-----PheIleIleLeuGluAsnIlePheVal----- 66  
Qy 350 GGAGCGTCATCGCTCGGGAGCTGTTCTTCTGAATCTTTTGTCTTATCTGGCAGCT--- 406  
Db 67 -----LeuLeuThrIleTrpLysThrLys 74  
Qy 407 -----CTACAGTATTGGATCGCCACTCACCTTGTCTCAGGACTCGCTGGG 451  
Db 75 LysPheHisArgProMetTyrTyrPheIleGlyAsnLeuAlaLeuSerAspLeuLeuAla 94  
Qy 452 ATTGTTCTTGTATGCGGATGCGCGTAGATGCCAAATGTTCTTGTATTTCGAAGAATCCGA 511  
Db 95 -----GlyValAlaTyrThrAlaAsnLeuLeu----- 104  
Qy 512 GAGGAATTTTATTGTCTCAAAAGCTTAAAAAATCTGTAGAAAAAGGATATACCAAGCT 571  
Db 104 ----- 104  
Qy 572 TTTGAGCCATTTTGTGATTTCTAAGTCTGACTACAGTATTGGCTCAGCAGCTTCTTTTCTTC 631  
Db 105 -----SerGlyAlaThrThrTyrLysLeuThrProAlaGlnTrpPhe 118  
Qy 632 CTAGATACAGGCGCTATTAAAGGCTTGTGCTTTGACATTGATTTTAGGAATTTTCTTTCA 691  
Db 119 LeuArgGluGly-----SerMetPheValAlaLeuSerAlaSer 131  
Qy 692 ATGTTTACG-----GCTCTTTTTCATGACTAAATTTTCTTCATGCTGTGGATG--- 739  
Db 132 ValPheSerLeuLeuAlaIleAlaIleGluArgTyrIleThrMetLeuLysMetLysLeu 151  
Qy 740 ---AATAAGACCCCAACATACACAGTTGATATGATGAATAAGTTCTGTTGGGATAAAGCAT 796  
Db 152 HisGlySerAsnAsnPheArgLeuPheLeuLeu----- 163  
Qy 797 GATTTCTTGAGAGATGCAAAAACTTTTGGCTGTCTTCTTGAAGTGTGTTTCTTTTAGGT 856  
Db 164 -----IleSerAlaCys-----TrpValIleSer-----LeuIleLeuGly 175  
Qy 857 TGCCTTGTCTCTCGGGTTGGAGCCTGGAAATCCGTTTGGGAATGGATTTTAAAGGAGGG 916  
Db 176 -----GlyLeuProIleMetGlyTrpAsnCysIle----- 185  
Qy 917 TATGCTTTACCTTTAATCCAAAGAGCATGGCATGGCATGTTGCTCAAAATGCGTGGC 976  
Db 185 ----- 185  
Qy 977 AAAGTTGTGCATAAACTACAGGAAGCTGTCTTCTTCTAGAGACTTCCGTATTCAAACA 1036  
Db 186 -----SerAlaLeuSerSer----- 190  
Qy 1037 TTTGATCTTCAGAAAAAGATCAAAATCTATTTTAGTGATAAAGCTTTAAGCTATACTAAG 1096  
Db 191 -----CysSerThrValLeuProLeuTyrHisLysHisTyrIleLeuPheCysThrThr 208  
Qy 1097 CAGATACAGAGCTCTCTCTAAATAATTAACGATCATGAGCTGGCGTATTATGTTGGGATT--- 1153  
Db 209 valPheThrLeuLeuLeuLeuSerIleValIleLeu-----TyrCysArgIleTyr 225  
Qy 1154 ---GTTGTGAGAAACAGGCGCTAGATTCTCTACGGAAACTCTAAACGAAACGAAAAATTT 1210  
Db 226 SerLeuValArgThrArgSerArgLeu-----ThrPheArgLysAsn----- 240  
Qy 1211 TGGTCAAAGGTAACGACAAACTATCGAAGAAATGCGTTATCATGGCGACCATCGGGTTT 1270  
Db 241 -----IleSerLysAlaSerArgSerSerGluAsnValAlaLeu 253  
Qy 1271 TTAGGAGCTTTGGCAATCATCTTCG---CTCTATGTGAGTTTGGCTTTTGAATGGCAATAT 1327  
Db 254 LeuLysThrValIleIleValLeuSerValPheIleAlaCys-----TrpAlaPro 270

QY 1328 GCTTTCAGTCGCGATGCGCTTATTCATGACCTTTGGCTACCTGAGCCTCTGTTT 1387  
||| |||  
Db 271 Leuphelleuleuleuleu,leuaspvalGlyCysIysValIysIthrCysAspIleLeuphe 290  
QY 1388 ATAGCACATTTCTTTTGAMGAAATTCATATAGATTGGCAAGCCATGTGCTTATATG 1447  
||| |||  
Db 291 ArgAlaIeIutryrPheleu-----ValIeu 298  
QY 1448 ACTGTATTTGGGGATTCATTAACAAATACCTTTGATCATTTTTCATTCGTATTCGTGAAGAT 1507  
||| |||  
Db 299 AlAlaIeuleuanserGlyIthrAsnProIleIle----- 309  
QY 1508 CGCAACGGAACCTGTTTACCCCTATGACGTGTTTATGATATGATGCCCTTCAAAAGACG 1567  
||| |||  
Db 310 -----TyrThrIeuthrAsnIysGIuMetArgArgAla 320  
QY 1568 TTCAGCCGC 1576  
||| |||  
Db 321 PheIleArg 323  
RESULT 12  
US-08-852-824-18  
; Sequence 18, Application US/08852824C  
; Patent No. 6060272  
; GENERAL INFORMATION:  
; APPLICANT: Li et al.  
; TITLE OF INVENTION: Human G-Protein Coupled Receptors  
; FILE REFERENCE: 1488,1220000  
; CURRENT APPLICATION NUMBER: US/08/852,824C  
; CURRENT FILING DATE: 1997-05-04  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 381  
; TYPE: PRT  
; ORGANISM: genomic  
US-08-852-824-18  
Alignment Scores:  
Pred. No.: 0.00888 Length: 381  
Score: 101.00 Matches: 83  
Percent Similarity: 31.83% Conservative: 58  
Best Local Similarity: 18.74% Mismatches: 116  
Query Match: 3.05% Indels: 186  
DB: Gaps: 22  
US-09-868-987-1 (1-1864) x US-08-852-824-18 (1-381)  
QY 290 ATTCAGCATGCTGTGGTGGCAATGCTTATTTGATGAGCGTATATTATAGATT 349  
||| |||  
Db 53 IleleuIleCysCys-----PheIleIleleuGIuAsnIlePheVal----- 66  
QY 350 GGAGCGCATGCTGCGAGAGCTGTTCTTCGATCTTTGCTTAACTTCGAGACT-- 406  
||| |||  
Db 67 -----LeuIeuthrIleItrIlyPheIthrIlys 74  
QY 407 -----CTACAGATTTTGGATGCGCCACTCACCTTGTCAGACTCGCTGGG 451  
||| |||  
Db 75 IysPheIthrIserPrometTyrPheIleGIuAsnIleAlaIeuserAspIleuAla 94  
QY 452 ATTGTTCTGCTATGGGATGCGCTAGATGCAATGTTCTTGATTCGAAGAATCGGA 511  
||| |||  
Db 95 -----GlyValAlaIytrIthrAlaAsnIleuLeu----- 104  
QY 512 GAGGAATTTTATGTCTCAAAAGCTTAAAAATCTGAGAAAAAGATATACCAAGGCT 571  
||| |||  
Db 104 ----- 104  
QY 572 TTGGAGCCATTTTGTATCTTAACCTACACAGATTTGGCTCAGCACTTCTTTCTTC 631  
||| |||  
Db 105 -----SerGIyAlaIthrIthrIlyIysIeuthrProAlaGIthrPhe 118  
QY 632 CTAGATACAGGCGCTTATTAAGGGTTTGCTTGCATTTGATTTTAAAGAAATTTCTCTTCA 691

Db 119 LeuArgGIuGIy-----SerMetPheValAlaIeuserAlaSer 131  
QY 692 ATGTTTACG-----GCTCTTTCATGACTAAATTTTCTTCATGCTGTGGATG----- 739  
||| |||  
Db 132 ValPheSerIeuleuAlaIleAlaIleGIuArgTyrIlethrMetIeulysMetIysIeu 151  
QY 740 ----AATPAGCCCAACATACACAGTTGCATATGCATGAATTAAGTTCGTGGGGAATAAAGCAT 796  
||| |||  
Db 152 HisAsnGIySerAsnAsnPheArgIeuleuLeu----- 163  
QY 797 GATTTCTTGAGAGATGCAAAAACTTTGGGCTGTGTTCTGGAAGTGTGTTTCTTTAGGT 856  
||| |||  
Db 164 -----IleSerAlaCys-----TrpValIleSer-----LeuIleuGIy 175  
QY 857 TGGGTGCTCTCGGCTTTGGAGCTTGAATTCGTTTGGGAATGATTTTAAAGAGGG 916  
||| |||  
Db 176 -----GlyLeuProIleMetGIyTrpAsnCysIle----- 185  
QY 917 TATGCTTTACCTTTAATCCAAAGACATGCGATGACGATGTGCTCAATGCGGTGC 976  
||| |||  
Db 185 ----- 185  
QY 977 AAAGTTGTGCATTAACCTACAGAGAGCTGCTTTCTTCTAGAGACTTCGCTATTAACA 1036  
||| |||  
Db 186 -----SerAlaIeuserSer----- 190  
QY 1037 TTTGGATCTTCAGAAAAGATCAAAATCTATTTGATATAGACTTAAAGCTTAACTATAG 1096  
||| |||  
Db 191 -----CysSerThrValIeuleuProIeuthrIlyHisIleTyrIleleuPheCysThr 208  
QY 1097 CAGATACAGACTCTCTCCAAAATTAACATATAGATAGCTGGGCTTATGCTGGGAT-- 1153  
||| |||  
Db 209 ValPheThrIeuleuIeuleuSerIleValIleIeu-----TyrCysArgIleTyr 225  
QY 1154 ----GTTGTCAGAAAAGAGCTTACATGTTCTCTACGAAACTCTPAAAGAAAGCAAAATTT 1210  
||| |||  
Db 226 SerIeuValArgIthrArgSerArgIleu-----ThrPheArgIysAsn----- 240  
QY 1211 TGGTCAAAAGGTAAAGCAAACTATGCAAGAAAATGCGTTATGAGCGACATCGGGCTT 1270  
||| |||  
Db 241 -----IleSerIyAlaSerArgSerGIuAsnValAlaIeu 253  
QY 1271 TTAGAGCTTTGGCATCATCTTG---CTCTATGTAGATTGGCTTGAATGGCAATAT 1327  
||| |||  
Db 254 LeuIysThrValIleIleValIeuserValPheIleAlaCys-----TyrAlaPro 270  
QY 1328 GCTTTCAGTCGCGATGCGCTTATTCATGACCTTTTGGCTACCTGTCAGCTTGT 1387  
||| |||  
Db 271 Leuphelleuleuleuleu,leuaspvalGlyCysIysValIysIthrCysAspIleLeuphe 290  
QY 1388 ATAGCACATTTCTTTTGAMGAAATTCATATAGATTGGCAAGCCATGTGCTTATATG 1447  
||| |||  
Db 291 ArgAlaIutryrPheleu-----ValIeu 298  
QY 1448 ACTGTATTTGGGGATTCATTAACAAATACCTTTGATCATTTTTCATTCGTATTCGTGAAGAT 1507  
||| |||  
Db 299 AlAlaIeuleuanserGlyIthrAsnProIleIle----- 309  
QY 1508 CGCAACGGAACCTGTTTACCCCTATGACGTGTTTATGATATGATGCCCTTCAAAAGACG 1567  
||| |||  
Db 310 -----TyrThrIeuthrAsnIysGIuMetArgArgAla 320  
QY 1568 TTCAGCCGC 1576  
||| |||  
Db 321 PheIleArg 323  
RESULT 13  
US-08-467-947A-28  
; Sequence 28, Application US/08467947A  
; Patent No. 6090575  
; GENERAL INFORMATION:  
; APPLICANT: LI, YI



APPLICANT: CAO, LIANG  
 APPLICANT: NI, JIAN  
 APPLICANT: GENTZ, REINER  
 APPLICANT: BULT, CAROL J.  
 APPLICANT: SUTTON III, GRANGER G.  
 APPLICANT: ROSEN, CRAIG A.  
 TITLE OF INVENTION: Polynucleotides Encoding Human G-Protein  
 TITLE OF INVENTION: Coupled Receptor GPR1  
 NUMBER OF SEQUENCES: 30  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: STERNER, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
 STREET: 1100 NEW YORK AVE., NW, SUITE 600  
 CITY: WASHINGTON  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20005

COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY DISK  
 COMPUTER: IBM PC COMPATIBLE  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PATENTIN RELEASE #1.0, VERSION #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/467,947A  
 FILING DATE: 06-JUN-1995  
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/04079  
 FILING DATE: 30-MAR-1995

ATTORNEY/AGENT INFORMATION:  
 NAME: STEFFE, ERIC K.

REGISTRATION NUMBER: 36,688  
 REFERENCE/DOCKET NUMBER: 1488.1140002/EKS/KLM  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-371-2600  
 TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 28:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 381 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: peptide

US-08-467-947A-28

Alignment Scores:  
 Pred. No.: 0.00888 Length: 381  
 Score: 101.00 Matches: 83  
 Percent Similarity: 31.83% Conservative: 58  
 Best Local Similarity: 18.74% Mismatches: 116  
 Query Match: 3.03% Indels: 186  
 DB: 3 Gaps: 22

US-09-868-987-1 (1-1864) X US-08-467-947A-28 (1-381)

QY 290 ATCTCAGCATGCTGGCTTGGCAAGCTTATGTTTGTAGAGCGGTATATPATAGATTT 349  
 Db |||||  
 53 ileLeuileCysCys-----PheileileLeuGluAsnIlePheVal----- 66  
 QY 350 GGAGGCGTCATCGCTCGGGAGCTGTTCTTCTGAATCTTTTGTCTATCTGGCAGCT--- 406  
 Db |||||  
 67 -----LeuLeuthrileTrpIlysthrLys 74  
 QY 407 -----CTACAGTATTGGATGCGCCACTCACCTGTCTCAGGACTCGCTGGG 451  
 Db |||||  
 75 LysPheHisArgProMetTyrTyrPheileGlyAsnLeuLeuSerAspLeuLeuLa 94  
 QY 452 ATTGTTCTGTATGGGATGCCGTAGATGCAATGTTTGTATTTCGAAGAATCCGA 511  
 Db |||||  
 95 -----GlyValAlaTyrThrAlaAsnLeuLeu----- 104  
 QY 512 GAGGAATTTTATTGTTCTCAAGTCTTAAAAAATCTGTAGAAAAAGGATATACCAAGCT 571  
 Db |||||  
 104 ----- 104

QY 572 TTGGAGCCATTTTGTATTCTAACTTCACTACAGTATTGGCTCAGCAGCTTCTTTTCTTC 631  
 Db |||||  
 105 -----SerGlyAlaThrThrTyrLysLeuThrProAlaGlnTrpPhe 118  
 QY 632 CTAGATACAGGCGCTATTAAAGGTTTGCCTTGACATTTGATTTAGGAATTTTCTCTTCA 691  
 Db |||||  
 119 LeuArgGluGly-----SerMetPheValAlaLeuSerAlaSer 131  
 QY 692 ATGTTTACG-----GCTCTTTTCATGACTAAATTTTCTTCTCATGCTGTGGATG----- 739  
 Db |||||  
 132 ValPheSerLeuLeuAlaIleAlaIleGluArgTyrIleThrMetLeuLysMetLysLeu 151  
 QY 740 ---AATAAGCCCAACATACACAGTTGCATATGATGAATAAGTTCGTGGGGAATAAGCAT 796  
 Db |||||  
 152 HisAsnGlySerAsnAsnPheArgLeuPheLeuLeu----- 163  
 QY 797 GATTCTTGTAGAGGATGCAAAAAAATTTTGGCTGTTTCTGGAAGTGTTTTCTTTTAGGT 856  
 Db |||||  
 164 -----IleSerAlaCys-----TrpValIleSer-----LeuileLeuGly 175  
 QY 857 TGCCTTGTCTCTCGGGTTTGGAGCGCTGGAATTCCTGTTTGGGAATGGATTTTAAAGGAGG 916  
 Db |||||  
 176 -----GlyLeuProIleMetGlyTrpAsnCysIle----- 185  
 QY 917 TATGCTTTTACCTTTTAATCCAAAGAGCATGGCATCAGCGATGTTGTCTCAAAATCGTGGC 976  
 Db |||||  
 185 ----- 185  
 QY 977 AAAGTTGTGCATAAATACAGGAAGCTGGTCTTCTTCTTAGAGAGCTTCCGTATTCAAACA 1036  
 Db |||||  
 186 -----SerAlaLeuSerSer----- 190  
 QY 1037 TTTGATCTTTCAGAAAAGATCAAAATCTATTTTAGTATAAGCTTTAAAGTACTATAAG 1096  
 Db |||||  
 191 -----CysSerThrValLeuProleuTyrHisLysHisTyrIleLeuPheCysThrThr 208  
 QY 1097 CAGATACGAGCCTCTCTCTAAATAATTAAACGATCATGAGCTGGCGTTTATTTGGGATT--- 1153  
 Db |||||  
 209 ValPheThrLeuLeuLeuSerIleValIleLeu-----TyrCysArgIleTyr 225  
 QY 1154 ---GTTGTCAGAAACAGCGCTAGATTTCTTACGGAACACTCTAAACGAAACGCAAAATTT 1210  
 Db |||||  
 226 SerLeuValArgThrArgSerArgLeu-----ThrPheArgLysAsn----- 240  
 QY 1211 TGGTCAAAGGTAAGCAGCAAACTATCGAAGAAATGCGTTTATCAGGCGACCATCGGCTT 1270  
 Db |||||  
 241 -----IleSerLysAlaSerArgSerSerGluAsnValAlaLeu 253  
 QY 1271 TTAGGAGCTTTGGCAATCATCTTG---CTCTATGTGAGTTTTCGCTTTTGAATGGCAATAT 1327  
 Db |||||  
 254 LeuLysThrValIleIleValLeuSerValPheIleAlaCys-----TrpAlaPro 270  
 QY 1328 GCTTTAGTCGCGTATGCGCTTTAATTCATGACCTTTTGGCTACTGTGCGAGCTTGTGTTT 1387  
 Db |||||  
 271 LeuPheileLeuLeuLeuAspValGlyCysLysValLysThrCysAspIleLeuPhe 290  
 QY 1388 ATACACATTTCTTTTTCAGAAAATTCAAATAGATTGCAAGCATTTGGTGTCTTAATG 1447  
 Db |||||  
 291 ArgAlaGluTyrPheLeu-----ValLeu 298  
 QY 1448 ACTGTTTGGGTTTTCATTAAACAATCTTTGATCATTTTGTATCGTATTTCGTGAAGAT 1507  
 Db |||||  
 299 AlaValLeuAsnSerGlyThrAsnProIlele----- 309  
 QY 1508 CGCCAAGCAACCTGTTTACCCTTATGATGTTTGTATTAATGATGCCCTTCAAAGACG 1567  
 Db |||||  
 310 -----TyrThrLeuThrAsnLysGluMetArgAla 320  
 QY 1568 TTCAGCCGC 1576  
 Db |||||  
 321 PheileArg 323





```

Db 415 -----AlaserValSerThrValLeuSerLeuLeu 425
Qy 1616 ATGCTTTGTTATAGCGGCTCTGTC-----TTAATTTGCAATTAAT 1663
Db 426 PheValLeuLeuGlnValAlaValLeuLysTyrTyrAlaPheAsnArgIleSerLeu 445
Qy 1664 ATGACCAATAGGATTTCTTGTGAACTTATCGCTCTTTATATGACCACTCTGTG 1723
Db 446 PheIleValLysLeuIleLeuGlyMetIleMetSerIleValAlaGlnThrValMet 465
Qy 1724 TTGTTATGTCGCGTAA 1741
Db 466 LeuAlaLeuProSerLys 471

RESULT 16
US-09-542-733-2
; Sequence 2, Application US/09542733
; Patent No. 6323333
; GENERAL INFORMATION:
; APPLICANT: Ping Tsui
; TITLE OF INVENTION: Mouse EDG1
; FILE REFERENCE: GP-70610
; CURRENT APPLICATION NUMBER: US/09/542,733
; EARLIER FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 382
; TYPE: PRT
; ORGANISM: MUS MUSCULUS
US-09-542-733-2

Alignment Scores:
Pred. No. 0.0225 Length: 382
Score: 97.50 Matches: 86
Percent Similarity: 31.57% Conservat: 57
Best Local Similarity: 18.98% Mismatches: 131
Query Match: 2.94% Indels: 179
Gaps: 22
DB: 4

US-09-868-987-1 (1-1864) x US-09-542-733-2 (1-382)
Qy 257 GATCTTGGAAAAAACAATGTACACAAGGCAATATCTCAGATGCTGTGGCTTGGCAATG 316
Db 42 AsphIsGlyIleLysLeuThrSerValPheIleLeuIleCysCys----- 57
Qy 317 CTATTGTTGATGAGCGGTATATTATGATTGGAGCGTCATCGCTTGGAGACTGT 376
Db 58 PheIleIleLeuGlnAsnIlePheVal----- 66
Qy 377 CTTCGATCTTTTGCTTATCGGCACT-----CTACAGATTTG 418
Db 67 -----LeuLeuThrIleTyrLysThrLysPheHisArgProMetTyrTyrPhe 83
Qy 419 GATCGCCACTCAGCTTGTCTAGACTCGCTGGAGATTGTTGCTANGGAGTGGCGCTA 478
Db 84 IleGlyAsnLeuAlaLeuSerAspLeuAla-----GlyValAlaLys 98
Qy 479 GATGCAATGTCTTGTATTTGAAAGAAATCCGAGAGAAATTTTATTTGTCTCAAGTCT 538
Db 99 ThrAlaAsnLeuLeu----- 104
Qy 539 AAAAAATGTGNAAGAAAGATATATCAAGGCTTTTGAGGCAATTTTGATCTTAAGTTG 598
Db 105 -----SerGlyAla 107
Qy 599 ACTACAGTATTGGCTCAGCACTTCTTTCTCTGATATACAGGCGCTATTAAAGGTTT 658
Db 108 ThrThrTyrLysLeuThrPheAlaGlnThrPheLeuAlaGlnGly----- 122

```

```

Qy 659 GCTTGACATGATTTAGGAATTTCTCTCAATGTTACG-----GCTCTTATG 712
Db 123 -----SerMetPheValAlaLeuSerAlaSerValPheSerLeuAlaIleAlaIle 140
Qy 713 ACTAAATTTTCTTCATGCTGTGATGAATTAAGACCAACATACAGTTGATATGATG 772
Db 141 GluArgTyrIleThrMetLeuLysMetLys-----LeuHisAsnGlySerAsnSer 158
Qy 773 AATTAATGCTGGGATTAAGCATGATTTCTTGAGAGATGCAAAAACCTTGGCTGT 832
Db 159 ArgSerPheLeu-----LeuIleSerAlaCys-----TyrValIle 170
Qy 833 TCTGAAGTGTCTTTCTTTAGGTGCTGTCGCTGCGGTTTGAGACCTGATTCCTGT 892
Db 171 Ser-----LeuIleLeuGly-----GlyLeuProIleMetGlyTyrPheCysIle 185
Qy 893 TTGGAAATGATTTTAAAGAGGGTATGCTTACCTTTAATCCAAAGACATGACATC 952
Db 185 ----- 185
Qy 953 AGCGATGTCCTCAATGCGTGGCAAGTTTGATTAATACAGAGAGCTGCTTCT 1012
Db 186 -----SerSerLeuSer 189
Qy 1013 TCTAGAGACTTCGATATTCAAACATTTGGATCTTCAGAAAAGATCAAAATCTATTAGT 1072
Db 190 Ser-----CysSerThrValLeuProLeuTyrHisLys 200
Qy 1073 GATTAAGCTTTAGCTATATACAGATACGAGCTCTCTCTTAATTAATACAGATCATG 1132
Db 201 HisTyrIleLeuPheCysThrThrValPheThrLeuLeuLeuSerIleValIleLeu 220
Qy 1133 AGCTGCGTTATGTGGAT-----GTTTCAGAAACAGCGCTGATTTCTCTACGGA 1186
Db 221 -----TyrCysArgIleTyrSerLeuValArgThrArgSerArgLeu----- 235
Qy 1187 AACTTAAGCAAGCAAAATTTGTCTAAAGGTAGCAGCAAGCAACTATCGAAGAAATG 1246
Db 236 ThrPheArgLysAsnIleSerLysAlaSerArgSerSerLys----- 250
Qy 1247 CGTTATACGAGCAATCGGCTTTTATAGAGCTTTGGCAATCATCTTG---CTCTATGTG 1303
Db 251 -----SerLeuAlaLeuLeuLysThrValIleIleValLeuSerValPheIle 266
Qy 1304 AGTTGGCGCTTGAATGGCAATATGCTTTGATGCGCTATATGCTTATATCATGACTT 1363
Db 267 AlaCys-----TyrAlaProLeuPheIleLeuLeuLeuAspValGlyCysLys 283
Qy 1364 TTGGCTACCTGTGCAGCTGTGTTATAGACATTTCTTTTGAAGAAATTCAAATAGAT 1423
Db 284 AlaLysThrCysAspIleLeuTyrLysAlaGlnTyrPheLeu----- 297
Qy 1424 TTGCAMGCCATGTGCTTAAATGACTGATTTGAGGATTCATTAACATTAATCTTGATC 1483
Db 298 -----ValLeuAlaValLeuAsnSerGlyThrAsnProIleIle--- 310
Qy 1484 ATTTTGAATGATATGATGTAAGATCGCAAGCAAGCAACTGTTTACCCCTATGACATGTTTA 1543
Db 311 -----TyrThrLeu 313
Qy 1544 GTTAATGATCCCTTCAAAAGAGTTTCAGCGCAGCGTA 1582
Db 314 ThrAsnLysGlnMetArgAlaPheIleArgIleVal 326

RESULT 17
US-07-998-283B-8
; Sequence 8, Application US/07998289B
; Patent No. 6027876
; GENERAL INFORMATION:
; APPLICANT: Black, Bruce C
; APPLICANT: Taylor, Martin
; APPLICANT: Heckel, David G
; TITLE OF INVENTION: Method for Monitoring Pesticide

```

TITLE OF INVENTION: Resistance  
 NUMBER OF SEQUENCES: 40  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Darby & Darby PC  
 STREET: 805 Third Avenue  
 CITY: New York  
 STATE: New York  
 COUNTRY: US  
 ZIP: 10022  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/998,289B  
 FILING DATE: 30-DEC-1992  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Robinson, Joseph R  
 REGISTRATION NUMBER: 33,448  
 REFERENCE/DOCKET NUMBER: 0646/OA939  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-527-7700  
 TELEFAX: 212-753-6237  
 TELEX: 236687  
 INFORMATION FOR SEQ ID NO: 8:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1820 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-07-998-289B-8

Alignment Scores:		
Pred. No.:	0.0514	Length:
Score:	97.50	Matches:
Percent Similarity:	33.58%	Conservative:
Best Local Similarity:	17.92%	Mismatches:
Query Match:	2.94%	Indels:
DB:	3	Gaps:
		24
		1820

US-09-868-987-1 (1-1864) x US-07-998-289B-8 (1-1820)

Qy	287	ATTATCTCAGCARGCTGTGGCTTGGCAAGCTTATTGTTTGTAGTACGGCTATATTAGA	346
Db	1319	LeuThrAsnAlaTpcCysTrpLeuAspPheValIleValMetValSerLeuIleAsnPhe	1338
Qy	347	TTTGGAGGCGTCATCGCTTCGGGAGCTGCTCTTCGTAATCTTTTGCCTTATCTGGGCGACT	406
Db	1339	ValAlaSerLeuValGlyAlaGlyGlyIle-----GlnAla	1350
Qy	407	CTACAGTATTTGGATGGCGCACTCACCTTGTTCAGACTCGCTGGGATTTGTTCTTGCTATG	466
Db	1351	PheLysThrMetArgThrLeuAlaLeuArgProLeuArgAlaMetSerArgMetGln	1370
Qy	467	GGGATGGCGGTAGATGCAAAATGTTCTGTATTGCAAGAAATCCGAGAGGAATTTTATTG	526
Db	1371	GlyMetArgValValAsnAlaLeuVal-----	1380
Qy	527	TCTCAAAGCTTAAAAAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGCCATTTT	586
Db	1381	-----GlnAlaIleProSerIlePhe	1387
Qy	587	GATTCTAACTGACTACGATATTGGCCTCAGCACTCTTTTCTTCCATAGATACAGGCGCT	646
Db	1388	Asn-----ValLeuLeuValCysLeuIlePheTrpLeu-----	1398
Qy	647	ATTAAAGGGTTGCTTTGCATGTATTTAGGAATTTCTCTCAATGTTTACGGCTCTT	706
Db	1399	-----IlePheAlaIleMetGlyValGlnLeu-----	1407

DB 1693 LeuLeuPheAlaLeuAlaMetSerLeuProAlaLeuPheAsnIleCys----- 1708  
 QY 1706 ATGCACACACCTCTGTGTGTGTATGATC 1735  
 DB 1709 -----LeuLeuLeuPheLeuVal 1714

RESULT 18  
 US-08-808-793-23  
 ; Sequence 23, Application US/08808793  
 ; Patent No. 5858713  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Soderlund, David M.  
 ; APPLICANT: Ingles, Patricia J.  
 ; TITLE OF INVENTION: CALCIUM PERMEABLE INSECT SODIUM CHANNELS  
 ; TITLE OF INVENTION: ANT USE THEREOF  
 ; NUMBER OF SEQUENCES: 32  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
 ; STREET: Clinton Square, P.O. Box 1051  
 ; CITY: Rochester  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 14603  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/808,793  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/012,649  
 ; FILING DATE: 01-MAR-1996  
 ; CLASSIFICATION: 435  
 ; INFORMATION NUMBER: 435  
 ; NAME: Bitman, Susan J.  
 ; REGISTRATION NUMBER: 34,103  
 ; REFERENCE/DOCKET NUMBER: 19603/1062 (D-1906A)  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 716-263-1636  
 ; TELEFAX: 716-263-1600  
 ; INFORMATION FOR SEQ ID NO: 23:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2100 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-808-793-23

Alignment Scores:  
 Pred. No.: 0.0554 Length: 2100  
 Score: 97.50 Matches: 95  
 Percent Similarity: 33.58% Conservative: 83  
 Best Local Similarity: 17.92% Mismatches: 171  
 Query Match: 2.94% Indels: 181  
 DB: 2 Gaps: 24

US-09-868-987-1 (1-1864) x US-08-808-793-23 (1-2100)

QY 287 ATATCTCAGCATGCTGTG3CTGGCAATGCTTATGTTTGATGAGCGTATATATATAGA 346  
 DB 1332 LeuThrAsnAlaTrpCysTrpLeuAspPheValIleValMetValSerIleuIleAsnPhe 1351  
 QY 347 TTGGAGCGGCATCGCTTCCGGACGCTTCTCTGGAATCTTTGCTTATCTGGGACGCT 406

DB 1352 ValAlaSerLeuValGlyAlaGlyIle-----GlnAla 1363  
 QY 407 CTACAGTATTTGGATGGCGGCACTACCTTGTGCAGACTGCGGATGTTCTTGTCTATG 466  
 DB 1364 PheLysThrMetLysThrLeuArgAlaLeuArgProLeuArgAlaMetSerArgMetGln 1383  
 QY 467 GGGATGGCCGTAGATGACAAATGTTCTTGATTCGAAAGATCCGAGAGATTTTATATG 526  
 DB 1384 GlyMetArgValValValAsnAlaLeuVal----- 1393  
 QY 527 TCTCAAGTCTTAAATAATCTGTAGAAAAAGATATACCAAGCTTTGGACCATTTT 586  
 DB 1394 -----GlnAlaIleProSerIlePhe 1400  
 QY 587 GATCTTACTTACCTACAGTATTTGGCTTCAGACCTTCTTCTTCTTCTTCTTCTTCTT 646  
 DB 1401 Asn-----ValLeuLeuValCysLeuIlePheTrpLeu----- 1411  
 QY 647 ATTAAGGTTTGCTTTGACATTTAGAAATTTCTTCAATTTTACGCTCTT 706  
 DB 1412 -----IlePheAlaIleMetGlyValGlnLeu 1420  
 QY 707 TTCATGACTAAATTTTCTTCATGCTGTGATGATGATTAAGCCACATACAGTTGAT 766  
 DB 1421 PheAlaGlyLysThrPheLysCysGluAspMetAsnGlyThrIys----- 1435  
 QY 767 ATGATGAATTAAGTCTGGGATTAAGCATGATTTCTTGAGAGATGCAAAAACCTTGG 826  
 DB 1436 -----LeuSerHisGluIleIleProAsnArgAsn----- 1445  
 QY 827 GCTGTTTCTGAAGTGTCTTTCTTTAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 886  
 DB 1446 -----AlaCysGluSerGluAsnTrpThrTrpValAsn 1456  
 QY 887 TCCGTTTGGGAATGATTTTAAAGA---GGGATGCTTTTACCTTAAATCCAAAGAG 943  
 DB 1457 SerAlaMetAsnPheAspHisValGlyAsnAlaTrpLeuCysLeuPhe----- 1472  
 QY 944 CATGCAATCAGCATGTTGCTCAAAATGCGTGGCAAGTTGTGATTAACATACAGAAAGCT 1003  
 DB 1473 -----GlnValAlaThrPheLysGlyTrpIle-----GlnIleMetAsnAsp 1486  
 QY 1004 GGTCTTCTCTCAGACACTCCGATTTCAAAACATTTGATCTTCAGAAAAGATCAAAATC 1063  
 DB 1487 AlaIleAspSerArgGluValAlaLeuLysGlnProIle-----ArgGluThrAsnIle 1503  
 QY 1064 TATTTAGTGAATTAAGCTTTAAGCTTAACTAAGACATACAGACCTCTCTCTTAAATTA 1123  
 DB 1504 TyrMetTrpLeuTrpPheValPhePheIle-----IlePheGlySerPhePheThrLeu 1521  
 QY 1124 ACGATCATGAGCTGGCGTATTTGGAATTTGTGCAAAACAGGCGCTAGATTCTCTAC 1183  
 DB 1522 AsnLeu-----PheIleGlyValIleIleAspAsn-----Phe 1532  
 QY 1184 GGAACCTTAACGAAACGCA----- 1204  
 DB 1533 AsnGluGlnLysLysLysAlaGlyLysSerLeuGluMetCysPheMetThrGluAspGlnLys 1552  
 QY 1205 AAATTTTGTCAAGATTAAGCAAGCAAACTATCGAAGAA----- 1243  
 DB 1553 LysTrpTrpSerAlaMetLysLysMetGlySerLysLysProLeuLysAlaIleProArg 1572  
 QY 1244 -----ATGGCTATACGGGACCATCGGCTTTTGAAGACT-----TTGGCA 1285  
 DB 1573 ProArgTrpArgProGlnAlaIleValPheGluIleValThrAspLysLysPheAspIle 1592  
 QY 1286 ATCATCTGCTCATGAGATTTG-----CCGTTTGAATGG 1321  
 DB 1593 IleIleMetLeuPheIleGlyLeuAsnMetPheThrMetThrLeuAspArgTrpAspAla 1612  
 QY 1322 CAATATGCTTCAAGTCCGATGCGCTTAAATTCATGACCTTTTG----- 1366  
 DB 1613 SerAspThrTrpAsnAlaValLeuAspTrpLeuAsnAlaIlePheValIlePheSer 1632

```
QY 1367 GCTACCTGTCAGCTCTGTTTATAGCA-----CATTCTCTTTTGAAGAAATTCAA 1417
Db 1633 SerGluCysLeuLeuLeuLeuLeuPheAlaLeuArgTyrHisTyrPheIleGluProTyrPheAsn 1652
QY 1418 ATAGATTGTCAGCAGCCTGGTCTTTAATGACTGTATTGGGTATTCTTAAACAATACT 1477
Db 1653 Leu---PheAspValValValValValLeuSerIleLeuGly----- 1665
QY 1478 TTGATCATTTTGTGTCGATTGTCGTAAGATCGCCAAAGCAACCTGTTTACCCCTATGCAT 1537
Db 1665 ----- 1665
QY 1538 GTTTAGTTATGATGCCCTTCAAAAGAGCTTC-----AGCCGACGGTAATGACACAGCT 1594
Db 1666 LeuValLeuSerAspIleGluIleGluTyrPheValSerProThrLeuLeuArgValVal 1685
QY 1595 ACAACTCTATCAGTTCTGTTAATGCTTTTGTATTATAGCGGCTCTCT-----GTC 1645
Db 1686 ArgValAlaLysValGlyArgValLeuArgLeuValLysGlyAlaLysGlyIleArgThr 1705
QY 1646 TTTAATTTGTCATTATTATGACCATAGGATTTCTTCTAGGAACCTTTATCGTCTCTTTAT 1705
Db 1706 LeuLeuPheAlaLeuAlaMetSerLeuProAlaLeuPheAsnIleCys----- 1721
QY 1706 ATTGCACCACCTCTGTTGTTGTTTATGTC 1735
Db 1722 -----LeuLeuLeuPheLeuVal 1727

RESULT 19
US-08-772-512A-19
; Sequence 19, Application US/08772512A.
; Patent No. 6022705
; GENERAL INFORMATION:
; APPLICANT: Soderlund, David M.
; APPLICANT: Knipple, Douglas C.
; APPLICANT: Ingles, Patricia J.
; TITLE OF INVENTION: INSECT SODIUM CHANNELS FROM
; TITLE OF INVENTION: INSECTICIDE-SUSCEPTIBLE AND INSECTICIDE-RESISTANT HOUSE
; TITLE OF INVENTION: FLIES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/772,512A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,618
; FILING DATE: 01-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Braman, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 19603/601 (CRFD-1657)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-263-1636
; TELEFAX: 716-263-1600
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2100 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
```

## MOLECULE TYPE: protein

US-08-772-512A-19

```
Alignment Scores:
Pred. No.: 0.0554 Length: 2100
Score: 97.50 Matches: 95
Percent Similarity: 33.58% Conservatives: 83
Best Local Similarity: 17.92% Mismatches: 171
Query Match: 2.94% Indels: 181
DB: 3 Gaps: 24
```

US-09-868-987-1 (1-1864) x US-08-772-512A-19 (1-2100)

```
QY 287 ATTATCTCAGCATCTCTGGCTTGGCAATGCTTATTGTTTGGAGCGGTATATATAGA 346
Db 1332 LeuThrAsnAlaTyrCysTrpLeuAspPheValIleValMetValSerLeuIleAsnPhe 1351
QY 347 TTTGGAGCGCTCATCGCTTCGGGAGCTGTTCTTCTGTAATCTTTTGTCTATCTGGCAGCT 406
Db 1352 ValAlaSerLeuValGlyAlaGlyGlyIle-----GlnAla 1363
QY 407 CTACAGTATTTTGGATGCGCCACTCACCTTGTTCAGGACTCGCTGGGATTGTTCTTGCTATG 466
Db 1364 PheLysThrMetArgThrLeuArgAlaLeuArgProLeuArgAlaMetSerArgMetGln 1383
QY 467 GGGATGCGCTAGATGCAAAATGTTCTTGTATTCGAAAGAATCCGAGAGGAATTTTATTG 526
Db 1384 GlyMetArgValValValAsnAlaLeuVal----- 1393
QY 527 TCTCAAGTCTTAAATAAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGCCATTTT 586
Db 1394 -----GlnAlaIleProSerIlePhe 1400
QY 587 GATTCTAACTTGACTACAGATGTTGGCTTCAGCACTCTTTTCTCTCTAGATACAGGGCCT 646
Db 1401 Asn-----ValLeuLeuValCysLeuIlePheTrpLeu----- 1411
QY 647 ATTAAAGGTTTGTGTTGACATGATTTTAGGAATTTCTCTCAATGTTTACGGCTCTT 706
Db 1412 -----IlePheAlaIleMetGlyValGlnLeu 1420
QY 707 TTCATGACTAAATTTTCTTCATGCTGTGGATGATTAAGACCCCAACATACACATGTCAT 766
Db 1421 PheAlaGlyLysTyrPheLysCysGluAspMetAsnGlyThrLys----- 1435
QY 767 ATGATGAATAAGTTTCGTGGGATAAAGCATGATTTCTTGAGAGGATGCAAAAACCTTTGG 826
Db 1436 -----LeuSerHisGluIleIleProAsnArgAsn----- 1445
QY 827 GCTGTTTCTGGAAGTGTGTTTCTTTTAGTTTGGCTTCTCTCGGGTTTGGAGCCTGGAAT 886
Db 1446 -----AlaCysGluSerGluAsnTyrThrTrpValAsn 1456
QY 887 TCGTGTGGAATGGATTTTAAAGGA---GGGTATCGCTTTTACCTTTTATCCAAAAGAG 943
Db 1457 SerAlaMetAsnPheAspHisValGlyAsnAlaTyrLeuCysLeuPhe----- 1472
QY 944 CATGGCATCAGCGATGTTGCTCAAAATCGGTGCAAAAGTGTGCAATAAACTACAGGAAGCT 1003
Db 1473 -----GlnValAlaThrPheLysGlyTrpIle-----GlnIleMetAsnAsp 1486
QY 1004 GGTCTTTCTTAGAGACTTCCCGTATTTCAAAACATTTTGGATCTTTCAGAAAAGATCAAAATC 1063
Db 1487 AlaIleAspSerArgGluValAspLysGlnProIle-----ArgGluThrAsnIle 1503
QY 1064 TATTTTAGTGAATAAGCTTTAAGCTATATAAGCAGATACAGAGCTCTCTCTCTAAATTA 1123
Db 1504 TyrMetTyrLeuTyrPheValPhePheIle-----IlePheGlySerPhePheThrLeu 1521
QY 1124 ACGATCATGAGCTGGCGTTATTGTTGGGATTTGTTCTCAGAAAACAGCGCTTAGATTCTCTAC 1183
Db 1522 AsnLeu-----PheIleGlyValIleIleAspAsn-----Phe 1532
```

```

QY 1184 GGAACCTTAACGAACGCA----- 1204
Db 1533 ASnGLuGLnlyslslyslaglyGlySerleuGLuMetPheMetThrGLuAspGLnLys 1552
QY 1205 AAATTTGGTCAAGGTAGACGACCAACTATCGAAGAA----- 1243
Db 1553 LysTrYrYserAlaMetLysLysMetGlySerLysLysProLeuLysAlaIleProArg 1572
QY 1244 -----ATGCGTTATCAAGCGACCATCGCGCTTTTAAAGACT-----TTGGCA 1285
Db 1573 ProArgTrpArgProGlnAlaIleValPheGluIleValThrAspLysLysPheAspIle 1592
QY 1286 ATCATCTTGGCTTATGTGANGTTG-----CGCTTTGANGG 1321
Db 1593 IleIleMetLeuPheIleGlyLeuAsnMetPheThrMetThrLeuAspArgTYrAspAla 1612
QY 1322 CAATATGCTTACAGTGGCGGCTTATGACCTTATGACCTTTTG----- 1366
Db 1613 SerAspThrTYrAsnAlaValLeuAspTYrLeuAsnAlaIlePheValValIlePheSer 1632
QY 1367 GCTACCTGTGACGCTGTGTTATAGCA-----CATTTCTTTTGAAGAAATTCAA 1417
Db 1633 SerGluSerLeuLeuLysIlePheAlaLeuArgTYrHisTYrPheIleGluProTrpAsn 1652
QY 1418 ATGATTTGGCAAGCCATTGCTGCTTAAATGACTGTATTGGGCTATTCAATTAACAATACT 1477
Db 1653 Leu---PheAspValIValValIValIleLeuSerIleLeuGly----- 1665
QY 1478 TTGATCATTTTGTATCGTATTCGTGAAGATCGCCAAAGCAACTGTTTACCCCTATGCAT 1537
Db 1665 ----- 1665
QY 1538 GTTTTAGTTAATGATGCCCTTCAAAAGACGTTC--AGCCGACGCGTAAATGACACAGCT 1594
Db 1666 LeuValLeuSerAspIleIleGluLysTYrPheValSerProThrLeuLeuArgValVal 1685
QY 1595 ACAACTATACGATTGTTGTTATGCTTTTGTATTAGCGGCTCCTC-----GTC 1645
Db 1686 ArgValAlaLysValGlyTYrGValLeuArgLeuValLysGlyAlaLysGlyIleArgThr 1705
QY 1646 TTTAATTTGATTTATTTATACCATAGGATTCCTTACAGAACTTATGCTCTTTAT 1705
Db 1706 LeuLeuPheAlaLeuAlaMetSerLeuProAlaLeuPheAsnIleCys----- 1721
QY 1706 ATTGCACCACCTCTGTGTGTTTATGCTC 1735
Db 1722 -----LeuLeuLeuPheLeuVal 1727

```

```

RESULT 20
US-09-134-001C-2912
: Sequence 2912, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lym Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: PRIOR FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 2912
: LENGTH: 400
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2912

```

```

Alignment Scores: 0.0263 Length: 400
Pred. No.: 97.00 Matches: 95
Score:

```

```

Percent Similarity: 35.71% Conservative: 90
Best Local Similarity: 18.34% Mismatches: 151
Query Match: 2.93% Gaps: 182
DB: 4 Gaps: 24
US-09-868-987-1 (1-1864) x US-09-134-001C-2912 (1-400)
QY 269 AATCAATGTAACACAGGCAATATCTACAGATGCTGCTGGCAATGCTTATGTTTG 328
Db 11 LysGluLeuLeuLeuIleMetIle-----AlaMetLeuIleValMet 24
QY 329 ATGAGCGTATATTATGATTTGAGGCGTCATCGCTTCGGAGACGTGCTTCTGAACTCT 388
Db 25 SerIleSerArgPheAlaIleValTYrThrProIle----- 34
QY 389 TTGCTTATCTGGGACCTCTACAGATTTGATGCGCCACTCACTTGCAGACCTGCT 448
Db 35 -----LeuProPheMetGlnGlnAspThrSerMetAspAngGln 48
QY 449 GGGATTGTTCTGCT-----ATGGGATGCGGTAGATCCAAATGTTCTGTA 496
Db 49 AlaGlyPheLeuAlaThrTrpAsnTYrLeuGlyTYrLeuValGlyAlaValIleProIle 68
QY 497 TTC-----GAAAGATCCGAGAGAAATTTTATGTTGCTCAAACTTTAA 541
Db 69 PheTYrIleTYrLysTYrLysValIleAspLeuLysIleTYrLeuAlaLeuAsnIleIle 88
QY 542 AATCTGTAGAAAAGATATACCAAGGCTTT-----GAGCCATTTTGTATCTAAC 595
Db 89 SerThrIleLeuMetGlyPheThrGlnGlnTYrLeuIleTrpSerIlePhe-----Arg 106
QY 596 TTGACTACAGTATTTGCGCTCAGACCTTCTTTCTTCTTCTGATACAGG----- 643
Db 107 LeuIleSerGlyIleThrSerGlyThrValPheValLeuThrSerAsnIleValLeuAsp 126
QY 644 CTTATTAAGGCTTTCCTTGAATGATTTTGAATTT-----TTCTCTCAATGTTTACG 700
Db 127 SerLeuLysLysIleGlnArgGlnGlyIleSerGlyValLeuTYrSerAlaValGlyLeu 146
QY 701 GCTCTTTTACGACTAAATTTTCTTCATGCTGTGATGATGAATTAAGCCAACTACACAG 760
Db 147 GlyIlePheLeuSerSerIleTYrLeuTYrPhePheAlaSer----- 160
QY 761 TTGCAATATGATGAATAAGTTCGTGGGATTAACAGATTTCTTGAGAGATGCAGAAA 820
Db 161 -----ValAspSerTrpLysMet 166
QY 821 CTTTGGGCTGTTCTGGAAGTGTCTTTCTT-----TTAGTTGCTGCTCTCGGGTTTGA 877
Db 167 ValTrpPheValLeuGlyIleValSerLeuValLeuGlyLeuIleValLeuValPheMet 186
QY 878 GCTGGAATTCGTTTGGGAATGATTTTAAAGAGGATGATCCCTTACCTTTAATCCA 937
Db 187 LysGluAsnThrAsnThrIleIleAsnPheIle----- 197
QY 938 AAGAGCATGCACTACAGCATGTGCTCAAAATCGTGCGCAAAATGTGCACTAACTACAG 997
Db 198 -----GluGluLysMetLysIleAsnAsnAsgLeu--- 207
QY 998 GAAGCTGCTTCTTCTTACAGACTCCGATTCAAACATTTGATCTTCAAGAAAGATC 1057
Db 208 -----AsnLysLysPhe 211
QY 1058 AAATCTATTTAGTATTAAGCTTTAAGCTAT----- 1090
Db 212 ValIlePhePheSer-----IleAlaTYrPheGluGlyAlaGlyTYrIleIle 228
QY 1091 -----ACTAAGCAATACAGACCTCTCTCCCTAAATTAACG 1126
Db 229 ThrGlyThrPheLeuValAlaIleValLysSerIle---ProSerValAlaGlyTYrAla 247
QY 1127 ATCATGAGCTGGCGTTATTTGGGATTTGTGTCAAGAAAGGCGCTAGATTTCTTACGGA 1186

```





```

APPLICATION NUMBER: US 60/005,498
FILING DATE: 04-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: UTSD:450
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1090 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-726-214-4

Alignment Scores:
Pred. No.: 0.0447 Length: 1090
Score: 97.00 Matches: 54
Percent Similarity: 37.98% Conservative: 44
Best Local Similarity: 20.93% Mismatches: 102
Query Match: 2.93% Indels: 59
Gaps: 11

US-09-868-987-1 (1-1864) x US-08-726-214-4 (1-1090)
QY 237 GTGAGAGAGCATCTCTTGATCTTGGGAAAAAC-----AATGTACACAGGCAATTA 290
Db 575 Ileglnaaglieserleuleuphetyrasmlyasnlleglulysglutyrargalathr 594
QY 291 TCT-----CAGCATGCTGTGGCTTGGCAATGCTTATTGTTGTGAGCGTAT 338
Db 595 AlaleuprolaaphelystyrvalthrCysAlaCysleuilepheleucysilephe 614
QY 339 ATTATAGATTG-----GAGGCGCA 359
Db 615 IilevalgnilleuvalleuprolysrleuileuglypheserPheglYAlaIa 634
QY 360 TCGCTTCGGAGAGCTGTCTTCTGAACTTTTGTCTTATCTGGGAGCTCTACAGTATTGG 419
Db 635 PheleuserleuilepheleuleuphevalCysphealaglylnleuileugln----- 652
QY 420 ATGGCGCATCACCCTTGCAGACTCGCTGGAGATTGTTCTGCTATG----- 467
Db 653 -----CyseryllysAlaserThrserleuMetIrpLeuLysSer 667
QY 468 -----GATGGCGCA-----GATGCAAT 487
Db 668 SerglyleileAlaAsnArgProIrpProArgIleSerleuThrIleValThrThraIa 687
QY 488 GTTCTTGATTGAAAGATCCGAGAGAAATTTTATTTGCTCAAGCTTTAAAAATCT 547
Db 688 IileleuThrMetAlaValPheAsnMetPhepheleuserAsnserGlnIuThrThr 707
QY 548 GTAGAAAAAGATATACCAAGGCTTTGGAGCCATTTTGTATTCACTTACATACAGTA 607
Db 708 LeuproThraIaAsnThrserAsnAlaAsnValSerValProAspAsnGlnAlaSerIle 727
QY 608 TTGGCCTACAGACTTTCTTCTCTAGATACAGGCGCTATTAAAGGTTTGCTTGACA 667
Db 728 LeuhIsAlaArgAsnLeuphePheleuproIyr-----PheIleTyser 742
QY 668 TTGATTTAGGAATTTCTCT--TCAATGTTTACGGCTCTT--TTCATGACTTAATTT 721
Db 743 CyslleleuglyleuilesserCysSerValPheleuAgtValAsrTyGlnleuLysMet 762
QY 722 TTCTTCATGCTG-----TGATGATTAAGACCCCAACCATACACAGTTCAT 766
Db 763 LeuileuMetValAlaValAlaValGlyTyraThrIleLeuLysThrIleAlaIaHis 782
QY 767 ATGATGAATAGTTGCTGG3GATA-----AAGCATGATTTCTTACAGAGATGCAAA 817

```

```

Db 783 ValLeuAspAlaIaThrSerGlnValLeuPheGlnArgProGlyIleTrpLysAspLeuLys 802
QY 818 AAACCTTGCGCTCTTTCGAAAGTGTCTTTTATGTTGCTGCTCTCGGG 871
Db 803 ThimelGlySerValSerleuSerIlePhePheIleThrleuValleuGly 820

RESULT 23
PCT-US95-11808-3
Sequence 3, Application PC/TUS9511808
GENERAL INFORMATION:
APPLICANT: Iyengar, Srinivas Ravi V.
TITLE OF INVENTION: MUTANT ACTIVATED GALPHA AND ADENYLYL
TITLE OF INVENTION: CYCLASE 2 FOR USE AS THERAPEUTIC AGENTS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Brumbaugh, Graves, Donohue and
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.
ZIP: 10112-0228
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/11808
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/307,896
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Richard S.
REGISTRATION NUMBER: 26,154
REFERENCE/DOCKET NUMBER: 29970 165/28755
TELEPHONE: (212) 408-2500
TELEFAX: (212) 765-2519
TELEX: 650 6111063
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1090 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-11808-3

Alignment Scores:
Pred. No.: 0.0447 Length: 1090
Score: 97.00 Matches: 54
Percent Similarity: 37.98% Conservative: 44
Best Local Similarity: 20.93% Mismatches: 102
Query Match: 2.93% Indels: 59
Gaps: 11

US-09-868-987-1 (1-1864) x PCT-US95-11808-3 (1-1090)
QY 237 GTGAGAGAGCATCTCTTGATCTTGGGAAAAAC-----AATGTACACAGGCAATTA 290
Db 575 Ileglnaaglieserleuleuphetyrasmlyasnlleglulysglutyrargalathr 594
QY 291 TCT-----CAGCATGCTGTGGCTTGGCAATGCTTATTGTTGTGAGCGTAT 338
Db 595 AlaleuprolaaphelystyrvalthrCysAlaCysleuilepheleucysilephe 614
QY 339 ATTATAGATTG-----GAGGCGCA 359
Db 615 IilevalgnilleuvalleuprolysrleuileuglypheserPheglYAlaIa 634

```

```
Qy 360 TCGCTTCGGGAGCTGTTCTTCTGAATCTTTTGGCTTATCTGGGCACTCTACAGTATTTGG 419
Db 635 PheLeuSerLeuIlePheIleLeuPheValCysPheAlaGlyGlnLeuLeuGln----- 652
Qy 420 ATGCGCCACTCACCTGTGTGTCAGGACTCGCTGGGATTGTTCTTGTCTATGG----- 467
Db 653 -----CysSerIlysIysAlaSerThrSerLeuMetTrpLeuLeuIysSer 667
Qy 468 -----GGATGGCCGTA-----GATGCAAAAT 487
Db 668 SerGlyIleIleAlaAsnArgProTrpProArgIleSerLeuThrIleValThrAla 687
Qy 488 GTTCTTGATTGCAAGAAGATCCGACAGAAATTTTATGCTCTCAAGTCTTAAAAATCT 547
Db 688 IleIleLeuThrMetAlaValPheAsnMetPhePheLeuSerAsnSerGluGluThrThr 707
Qy 548 GTAGAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTATCTTAACTTGACTACAGTA 607
Db 708 LeuProThrAlaAsnThrSerAsnAlaAsnValSerValProAspAsnGlnAlaSerIle 727
Qy 608 TTGGCCTCAGCATTCTTTCTTCTAGATACAGGCGCTATTAAAGGTTTGTCTTTGACA 667
Db 728 LeuHisAlaArgAsnLeuPhePheLeuProTyr-----PheIleTyrSer 742
Qy 668 TTGATTTAGGAATTTCTCT---TCAATGTTTACGGCTCTT---TTCATGACTAAATTT 721
Db 743 CysIleLeuGlyLeuIleSerCysSerValPheLeuArgValAsnTyrGluLeuIysMet 762
Qy 722 TTCCTCATGCTG-----TGGATGAATAAGCCACACATACAGTTGCTAT 766
Db 763 LeuIleMetMetValAlaLeuValGlyTyrAsnThrIleLeuLeuHisThrHisAlaHis 782
Qy 767 ATGATGAATAAGTTCGTGGGATA-----AAGCATGATTTCTTGAGAGGATGCAAA 817
Db 783 ValLeuAspAlaTyrSerGlnValLeuPheGlnArgProGlyIleTrpIysAspLeuIys 802
Qy 818 AACTTTGGGCTGTTCTGGAGTCTTTTCTTTAGTGGTGGCTGCTCGGG 871
Db 803 ThrMetGlySerValSerLeuSerIlePhePheIleThrLeuLeuValLeuGly 820
RESULT 24
US-09-262-477-2
; Sequence 2, Application US/09262477A
; Patent No. 6423508
; GENERAL INFORMATION:
; APPLICANT: George Livi
; APPLICANT: Derk Bergema
; APPLICANT: Jeffrey Stadel
; APPLICANT: Winnie Chan
; APPLICANT: Randall Johnson
; APPLICANT: Shelagh Wilson
; APPLICANT: Jon Chambers
; APPLICANT: Phillippe Robert
; APPLICANT: Nassirah Khandoudi
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDE SEQUENCES OF HUMAN EDG-1C AND
; TITLE OF INVENTION: METHODS OF SCREENING FOR AGONISTS AND ANTAGONISTS FOR HUMAN
; TITLE OF INVENTION: EDG-1C RECEPTOR AND SPHINGOSINE 1-PHOSPHATE LIGAND AND
; TITLE OF INVENTION: INTERACTION THEREOF
; FILE REFERENCE: GP50006
; CURRENT APPLICATION NUMBER: US/09/262,477A
; CURRENT FILING DATE: 1999-03-04
; EARLIER APPLICATION NUMBER: 60/077,369
; EARLIER FILING DATE: 1998-03-09
; EARLIER APPLICATION NUMBER: 60/087,102
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Human
US-09-262-477-2
```

```
Alignment Scores:
Pred. No.: 0.0293 Length: 382
Score: 96.50 Matches: 83
Percent Similarity: 32.05% Conservative: 59
Best Local Similarity: 18.74% Mismatches: 116
Query Match: 2.91% Indels: 185
DB: 4 Gaps: 22

US-09-868-987-1 (1-1864) x US-09-262-477-2 (1-382)
Qy 290 ATCTCAGCATGCTGCTGGCAATCTTATCTTGTATGAGCGGTATATTATAGATTT 349
Db 53 IleuIleCysCys-----PheIleIleLeuGluAsnIlePheVal----- 66
Qy 350 GGAGGCGTCATCGCTTCGGGAGCTGTTCTTCTGAATCTTTTGTCTTATCTGGGAGCT--- 406
Db 67 -----LeuLeuThrIleTrpLysThrLys 74
Qy 407 -----CTACAGTATTGGATGCGGCACCTCACCTTGTACGAGCTCGCTGGG 451
Db 75 LysPheHisArgProMetTyrTyrPheIleGlyAsnLeuAlaLeuSerAspLeuLeuAla 94
Qy 452 ATTGTTCTGTATGGGATGGCGGTAGATGCAAAATGTTCTTGTATTTCGAAAGAAATCCGA 511
Db 95 -----GlyValAlaTyrThrAlaAsnLeuLeu----- 104
Qy 512 GAGGAATTTTATTGTCTCAAAAGTCTTAAAAAATCTGTAGAAAAAGGATATACCAAGGCT 571
Db 104 ----- 104
Qy 572 TTGGAGCCATTTTGTATCTTAACCTGACTACATATTTGGCTCAGCAGCTTCTTTTCTTC 631
Db 105 -----SerGlyAlaThrThrTyrLysLeuThrProAlaGlnTrpPhe 118
Qy 632 CTAGATACAGGCGCTATTAAAGGTTTTCCTTGGACATTGATTTTAGGAATTTTCTCTTCA 691
Db 119 LeuArgGluGly-----SerMetPheValAlaLeuSerAlaSer 131
Qy 692 ATGTTTACG-----GCTCTTTTCATGACTAAAATTTTCTTCATGCTGTGGATG----- 739
Db 132 ValPheSerLeuLeuAlaIleAlaIleGluArgTyrIleThrMetLeuLeuMetLysLeu 151
Qy 740 ---AATAAGACCCCAACATACACAGTTCGATATGATGAATTAAGTTCGTGGGGAATAAGCAT 796
Db 152 HisAsnGlySerAsnAsnPheArgLeuPheLeuLeu----- 163
Qy 797 GATTCTTGAGAGGATGCAAAAAAATTTGGGCTGTTTCTGGAAGTGTTTTCTTTTAGGT 856
Db 164 -----IleSerAlaCys-----TrpValIleSer-----LeuIleLeuGly 175
Qy 857 TGCCTGCTCTCGGGTTTGGAGCCTCGAATTCCTGTTTGGGAATGGATTTTAAAGGAGGG 916
Db 176 -----GlyLeuProIleMetGlyTrpAsnCysIle----- 185
Qy 917 TATGCTTTTACCTTTTAATCCAAAAGACATGGCATCAGCGATGTTGTCAAATCGTGGC 976
Db 185 ----- 185
Qy 977 AAAGTTGTGCATAAATACAGGAAGCTGGTCTTCTTCTAGAGACTTCCCGTATTCAAACA 1036
Db 186 -----SerAlaLeuSerSer----- 190
Qy 1037 TTTGGATCTTCCAAAAGATCAAAATCTATTATTTAGTATAAGCTTTTAAGCTATATAAG 1096
Db 191 -----CysSerThrValLeuProLeuTyrHisLysHisTyrIleLeuPheCysThrThr 208
Qy 1097 CAGATACAGGCGCTCTCTCTTAAATAATTAAACGATCATGAGCTGGCGGTATTGTGGGAT--- 1153
Db 209 ValPheThrLeuLeuLeuSerIleValIleLeu-----TyrCysArgIleTyr 225
Qy 1154 ---GTTGTACAGAACAGGCGCTAGATTTCTCTACGGAACCTCTAAACGAAACCGCAAAATTT 1210
Db 226 SerLeuValArgThrArgSerArgArgLeu-----ThrPheArgLysAsnIleSerLys 243
```





QY 1049 GAAAGATCAAAATCTATTAGTATAAGCTTAACTACTAGACGATACGAGC 1108  
Db 301 -----ArgProLeuLeuCyStrpArgInglYlsGly 311  
QY 1109 TCT-----CTCTAAATTAACATCATGACG 1135  
Db 312 AlathrGlyArgArgGlyValYasnProGlyHisArgLeuLeuProLeuAArgSerSer 331  
QY 1136 TGGCGTTATTTGGGATTSTTGTCAAGAAACAGGCGCTAGATTCTCTACGAAACTCT 1192  
Db 332 SerLeuGluArgGlyLeuHisMetProThrSerProThrPheLeuGluGlyYasnThr 350

RESULT 27  
US-08-760-936-2  
Sequence 2, Application US/08760936  
Patent No. 5856443  
GENERAL INFORMATION:  
APPLICANT: MacLennan, A. John  
TITLE OF INVENTION: Molecular Cloning and Expression of  
NUMBER OF INVENTION: G-Protein Coupled Receptors  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: US  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/760,936  
FILING DATE: December 6, 1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Doran R.  
REGISTRATION NUMBER: 38,261  
REFERENCE/DOCKET NUMBER: MAC-100C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 352-375-8100  
TELEFAX: 352-372-5800  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 352 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-760-936-2

Alignment Scores:  
Pred. No.: 0.0545 Length: 352  
Score: 94.06 Matches: 77  
Percent Similarity: 37.73% Conservative: 66  
Best Local Similarity: 20.32% Mismatches: 136  
Query Match: 2.84 Indels: 100  
DB: 2 Gaps: 16

US-09-868-987-1 (1-1864) x US-08-760-936-2 (1-352)  
QY 215 TCTTTGTTCCGAGGTTTTCAGT---GAAGAGACGATCTTCTGATCTTGGGAAAAA 271  
Db 19 AenlyrThrLysGluThrLeuAspMetGlnGluThrProSer-----ArgLys 34  
QY 272 CAATGTACCAAGCATATATCTACAGCATCTGTGGCTTGGCAATG-----CTATATGTT 325  
Db 35 ValAlaSerAlaPheIleIleLeuLeuCySAlaIleValIleValGluAsnLeuVal 54  
QY 326 TTGAAGAGCGTA-----TATTATAGATTGGAGCGTC 358  
|||||

Db 55 LeuIleAlaValAlaArgAsnSerLysPheHisSerAlaMetLysLeuPheLeuGlyYasn 74  
QY 359 ATCGCTTGGGAGGTGTTCTTTCATCTTTGCTTATCTGGGAGCTTACAGTATTG 418  
Db 75 LeuAlaIaSerAspLeuLeuAlaGlyValAlaPheVal-----AlaAenThrLeuLeu 92  
QY 419 GATGGCGACACCTGCTGTC-----GGACTGGCT 448  
Db 93 SerGlyProValThrLeuSerLeuThrProLeuGlnThrPheAlaArgGluGlySerAla 112  
QY 449 GGGATTGTTCTGTATGGGATGGGATGCCGTAGATCAATGTTCTTGTATTCGAAAGATC 508  
Db 113 PheIleThrLeuSerAlaSer-----ValPheSer----- 122  
QY 509 CGAAGAGATTATTATGTTCTCAAGTCTTAAATAATCTGTAGAAAAAGATATACAG 568  
Db 123 -----LeuLeuAlaIleAlaIleGluArgGlnValAlaIleAlaValYals 138  
QY 569 GCTTTTGGAGCATTTTGTATCTTACTTACTACATCAAGTATTTGGCTCAGACCTTTTC 628  
Db 139 LeuYrGlySerAspLysSerCyAspMetLeuLeuIleGlyAlaSerTrpLeuIle 158  
QY 629 TTCCTAGATACAGG-----CTATTAAAGGTTT-----GCT 661  
Db 159 SerLeuIleLeuGlyGlyLeuProIleLeuGlyTrpAsnCySLeuAspHisLeuGluAla 178  
QY 662 TTGACATTGATTTTGAAGATTCTCTCAATGTTTACGCGCTCTTTTCATGACTAAATT 721  
Db 179 CysSerThrValLeuProLeuYrAlaLysHisLysYrValLeuCySValAlaThrIlePhe 198  
QY 722 TTTCTCATGCTGTGATGAATTAAGACCCACATACAGTTGCATATGATGAATTAAGTTC 781  
Db 199 SerValIleLeuLeuAlaIleValAlaLeuYrAlaGlyIleYrPheVal----- 215  
QY 782 GTGGGATTAAGCATGATTTCTTGAGAGATGCAAAATCTTGGGCTGTTCTGCAAGT 841  
Db 216 ValArgSerSerHisAlaAspValAlaGlyProGlnThrLeuAlaLeuLeuYThrVal 235  
QY 842 GTTTTCTTTTAGGTTGCGTTGCTCTCGGGTTTGGACCTGGAATTCGGTTTGGCAATG 901  
Db 236 ThrIleValLeuGlyValPheIleIleCySerTrpLeuProAlaPheSerIleLeuLeuLeu 255  
QY 902 GAT-----TTTAAAGAGGATAGCTTTACC 928  
Db 256 AspSerThrCySProValArgAlaCySProValLeuYrLysAlaHisLysPhePheAla 275  
QY 929 TTTATTCMAAAGACATGATGATGATGCTCAAAATGCTGCAAAAGTTGTGCAT 988  
Db 276 Phe-----AlaThrLeuAsnSerLeuLeuAsnProValIleTyr 288  
QY 989 AACTACAGAGCTGATCTTTCTTCTAGACACTTCGATTTCAACATTTGGATCTTCA 1048  
Db 289 ThrTrpArg-----SerArgAspLeuArgGluValLeu----- 300  
QY 1049 GAAAGATCAAAATCTATTATAGTATAAGCTTAACTACTACTAGACGATACGAGC 1108  
Db 301 -----ArgProLeuLeuCyStrpArgInglYlsGly 311  
QY 1109 TCT-----CTCTAAATTAACATCATGACG 1135  
Db 312 AlathrGlyArgArgGlyValYasnProGlyHisArgLeuLeuProLeuAArgSerSer 331  
QY 1136 TGGCGTTATTTGGGATTSTTGTCAAGAAACAGGCGCTAGATTCTCTACGAAACTCT 1192  
Db 332 SerLeuGluArgGlyLeuHisMetProThrSerProThrPheLeuGluGlyYasnThr 350

RESULT 28  
US-09-134-001C-5116  
Sequence 5116, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS







```
QY 1193 AACGAAACGCAAAATTTTGTCAAGGTAGCAGCAAACTATCGAAGAAATCGGTTAT 1252
Db 944 -----PheTrpGluGlnTyrMetThrLeuArgSerSerLeuAlaMetile 958
QY 1253 CAGCGCAACATCGGCTTTTAGGAGCTTTGGCAATCATC--TTGCTCTATGTGAGTTTG 1309
Db 959 LeuAlaCysVal---LeuLeuAlaLeuValLeuValSerLeuLeuLeuSerVal 977
QY 1310 CGCTTTGAATGCAATATGCTTTCAGTCCGTATGCGCTTTAAATTCATGACCTTTTGCT 1369
Db 978 -----Trp-----AlaAlaValLeuValLeuSerValLeuAla 989
QY 1370 ACCTGTGAGTCTGCTTTTATAGCACATTTCTTTTGAAGAAATTCATAGATTGCAA 1429
Db 990 SerLeuAlaGlnIlePheGlyAla-----MetThrLeuLeuGlyIleLysLeuSer 1006
QY 1430 GCCATTGCTGCTTTAAATGACTGTATGCGGTATTCATTAACAACTATCGAAGAAATCGGTTAT 1489
Db 1007 AlaIleProAlaValIleLeuLeuSerValGlyMet-----MetLeuCysPhe 1023
QY 1490 GATCGTATTCGTGAAGATCGCAAGCAACCTGTTTACCCCTATGCTATGTTTAAAT 1549
Db 1024 -----AsnValLeuIleSer 1028
QY 1550 GATGCCCTTCAAAGACGTTTCAGC-----CGCACGGTAATGACACAGCTTACA 1597
Db 1029 LeuGlyPheMetThrSerValGlyAsnArgGlnArgValGlnLeuSerMetGlnMet 1048
QY 1598 ACTGTATCAGTCTTTGTTAAATGCTTTTCTTTATAGCGCGCTCCTCTGCTTTT----- 1648
Db 1049 SerLeuGlyProLeuValHisGlyMetLeuThrSerGlyValAlaValPheMetLeuSer 1068
QY 1649 -----AATTTGCAATTTATATGACCATAGGATTTCTT 1681
Db 1069 ThrSerProPheGluPheValIleArgHisPheCysTrpLeuLeuValValLeuCys 1088
QY 1682 CTAGAACTTTATCGTCTCTTTTATATGACCACTCTGTTGTTTATGCTCGGTAAA 1741
Db 1125 ProLeuProValArgSerSerLys 1132
```

## RESULT 31

```
US-08-954-668-6
; Sequence 6, Application US/08954668
; Patent No. 6172200
; GENERAL INFORMATION:
; APPLICANT: SCOTT, MATHEW P
; APPLICANT: GOODRICH, LISA V
; APPLICANT: JOHNSON, RONALD L
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley, Hoag & Eliot
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/954,668
; FILING DATE: 20-Oct-1997
```

```
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36709
; REFERENCE/DOCKET NUMBER: SUV-003.06
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1285 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-954-668-6
Alignment Scores:
Pred. No.: 0.123 Length: 1285
Score: 93.50 Matches: 62
Percent Similarity: 42.91% Conservative: 53
Best Local Similarity: 23.13% Mismatches: 90
Query Match: 2.82% Indels: 63
DB: 4 Gaps: 14
US-09-868-987-1 (1-1864) x US-08-954-668-6 (1-1285)
QY 1082 TTAAGCTATATAGCAGATACGAGCCCTCTCTCTAAATTAACGATCATGAGCTGGCGT 1141
Db 908 LeuThrAspThrSerGlnIleLysThrLeuIleGlyHisIleArgAspLeuSerValLys 927
QY 1142 TATTGTGGGATTTGTCAGAAACAGCCCTAGA-----TTTCTCTACGGAACCTCT 1192
Db 928 TyrGluGlyPheGlyLeuProAsnTyrProSerGlyIleProPheIle----- 943
QY 1193 AACGAAACGCAAAATTTTGTCAAGGTAGCAGCAAACTATCGAAGAAATCGGTTAT 1252
Db 944 -----PheTrpGluGlnTyrMetThrLeuArgSerSerLeuAlaMetile 958
QY 1253 CAGCGCAACATCGGCTTTTAGGAGCTTTGGCAATCATC--TTGCTCTATGTGAGTTTG 1309
Db 959 LeuAlaCysVal---LeuLeuAlaLeuValLeuValSerLeuLeuLeuSerVal 977
QY 1310 CGCTTTGAATGCAATATGCTTTCAGTCCGTATGCGCTTTAAATTCATGACCTTTTGCT 1369
Db 978 -----Trp-----AlaAlaValLeuValLeuSerValLeuAla 989
QY 1370 ACCTGTGAGTCTGCTTTTATAGCACATTTCTTTTGAAGAAATTCATAGATTGCAA 1429
Db 990 SerLeuAlaGlnIlePheGlyAla-----MetThrLeuLeuGlyIleLysLeuSer 1006
QY 1430 GCCATTGCTGCTTTAAATGACTGTATGCGGTATTCATTAACAACTATCTTTGATCATTTT 1489
Db 1007 AlaIleProAlaValIleLeuLeuSerValGlyMet-----MetLeuCysPhe 1023
QY 1490 GATCGTATTCGTGAAGATCGCAAGCAACCTGTTTACCCCTATGCTATGTTTAAAT 1549
Db 1024 -----AsnValLeuIleSer 1028
QY 1550 GATGCCCTTCAAAGACGTTTCAGC-----CGCACGGTAATGACACAGCTTACA 1597
Db 1029 LeuGlyPheMetThrSerValGlyAsnArgGlnArgValGlnLeuSerMetGlnMet 1048
QY 1598 ACTGTATCAGTCTTTGTTAAATGCTTTTCTTTATAGCGCGCTCCTCTGCTTTT----- 1648
Db 1049 SerLeuGlyProLeuValHisGlyMetLeuThrSerGlyValAlaValPheMetLeuSer 1068
QY 1649 -----AATTTGCAATTTATATGACCATAGGATTTCTT 1681
Db 1069 ThrSerProPheGluPheValIleArgHisPheCysTrpLeuLeuValValLeuCys 1088
QY 1682 CTAGAACTTTATCGTCTCTTTTATATGACCACTCTGTTGTTTATGCTCGGTAAA 1741
Db 1125 ProLeuProValArgSerSerLys 1132
```

Db 1089 ValGlyAlaCysaenSerLeuValPhe---ProIleLeuLeuSerMetValGlyPro 1107  
 QY 1742 GAAATGCTCAAAATAAGTACCGTTAACTTAATCTTAACGTTAGCAATATAAAATCT 1801  
 Db 1108 Glu-----AlaGluLeuValProLeuGluHisProAspArgIleSerThr---ProSer 1124  
 QY 1802 CCTTGGGACTTGTAGTCCCAAGG 1825  
 Db 1125 ProLeuProValArgSerSerLys 1132

RESULT 32  
 US-08-918-658-6  
 ; Sequence 6, Application US/08918658  
 ; Patent No. 6423154  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SCOTT, MATHEW P  
 ; GOODRICH, LISA V  
 ; JOHNSON, RONALD L  
 ; TITLE OF INVENTION: Patched Genes and their Use  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
 ; STREET: Four Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: CA  
 ; COUNTRY: US  
 ; ZIP: 94111  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/918,658  
 ; FILING DATE: 22-Aug-1997  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/656,055  
 ; FILING DATE: <Unknown>  
 ; APPLICATION NUMBER: 08/540,406  
 ; FILING DATE: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rowland, Berttram I  
 ; REGISTRATION NUMBER: 20015  
 ; REFERENCE/DOCKET NUMBER: a60190-1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-781-1989  
 ; TELEFAX: 415-358-3249  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1285 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
 ; US-08-918-658-6

Alignment Scores:  
 Pred. No.: 0 123 Length: 1285  
 Score: 93.50 Matches: 62  
 Percent Similarity: 42.91% Conservative: 53  
 Best Local Similarity: 23.13% Mismatches: 90  
 Query Match: 2.82% Indels: 63  
 DB: 4 Gaps: 14

US-09-868-987-1 (1-1864) x US-08-918-658-6 (1-1285)  
 QY 1082 TTAAGCTACTAAGCAGATGAGACCTCTCTCTTAATAATTAACGATCATGAGCTGGCGT 1141  
 Db 908 LeuThrAspThrSerGlnIleLysThrLeuIleGlyHisIleArgAspLeuSerValLys 927  
 QY 1142 TATTGGGATTGTTGTCAAAAACAGGCTTAGA-----TTTCTCTACGAAACTCT 1132

Db 928 TyrGluGlyPheGlyLeuProAsnTyrProSerGlyIleProPheIle----- 943  
 QY 1193 AAACGAAACGCAAAATTTGGTCAAGGTAAGCAGCAAACTATGCAAGAAATGCGTTAT 1252  
 Db 944 -----PheTrpGluGlnTyrMetThrLeuArgSerSerLeuAlaMetIle 958  
 QY 1253 CAGCGACCATCGGGCTTTTATAGAGCTTTTGGCAATATC---TTGGCTATAGTAATTGG 1309  
 Db 959 LeuAlaCysVal---LeuLeuAlaAlaLeuValSerLeuLeuLeuSerVal 977  
 QY 1310 CGCTTGAATGCAATATGCTTTCAGTCCGTATGCGCTTAATTCATGACCTTTGGCT 1369  
 Db 978 -----Tyr-----AlaAlaValLeuValIleLeuSerValLeuAla 989  
 QY 1370 ACCTGTCAGCTCTGTATTATAGCATTCTTTTGAAGAAATTAAGATTGGCA 1429  
 Db 990 SerLeuAlaGlnIlePheGlyAla-----MetThrLeuGlyIleLysLeuSer 1006  
 QY 1430 GCGATTGCGCTTAAATGACTGATTTGGGATTCATTAAACAATACTTTGATCATTTT 1489  
 Db 1007 AlaIleProAlaValIleLeuIleLeuSerValGlyMet-----MetLeuCysPhe 1023  
 QY 1490 GATCGTATTGTGAAGATGCGCAAGCAACTGTATTACCCCTATGATGATTTGTTAAT 1549  
 Db 1024 -----AsnValLeuIleSer 1028  
 QY 1550 GATGCCCTTCAAAAGCGTTTCAGC-----CGCAGGTAATGACACAGCTTAC 1597  
 Db 1029 LeuGlyPheMetThrSerValGlyAsnArgGlnArgArgValGlnLeuSerMetGlnMet 1048  
 QY 1598 ACTCTATACGTTTGTATTATGCTTTTGTATAGGGGCTCCCTCTGCTTT----- 1648  
 Db 1049 SerLeuGlyProLeuValHisGlyMetLeuThrSerGlyValAlaValPheMetLeuSer 1068  
 QY 1649 -----AATTTGCAATTAATTAATGACCAATGGATTCTT 1681  
 Db 1069 ThrSerProPheGluPheValIleArgHisPheCysTrpLeuLeuValValLeuCys 1088  
 QY 1682 CTAGGAACCTTATTCGTCTCTTAATTAATGACACACCTGTGTGTTTATGTCGTAA 1741  
 Db 1089 ValGlyAlaCysaenSerLeuValPhe---ProIleLeuLeuSerMetValGlyPro 1107  
 QY 1742 GAAATGCTCAAAATAAGTACCGTTAACTTAATCTTAACGTTAGCAATATAAAATCT 1801  
 Db 1108 Glu-----AlaGluLeuValProLeuGluHisProAspArgIleSerThr---ProSer 1124  
 QY 1802 CCTTGGGACTTGTAGTCCCAAGG 1825  
 Db 1125 ProLeuProValArgSerSerLys 1132

RESULT 33  
 PCT-US95-13233-6  
 ; Sequence 6, Application PC/TUS9513233  
 ; GENERAL INFORMATION:  
 ; APPLICANT: THE BOARD OF TRUSTEES OF TEH LELAND STANFORD JUNIOR UNIVERSITY  
 ; TITLE OF INVENTION: Patched Genes and their Use  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
 ; STREET: Four Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: CA  
 ; COUNTRY: US  
 ; ZIP: 94111  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/13233  
 ; FILING DATE: 06-OCT-1990



QY 1550 GATGCCCTTCAAAAAGAGCTTCAGC-----CGCAGGTAATGACACAGCTGACA 1597  
 Db 1030 LeuGlyPheMetThSerValGlyAsnArgGlnArgValGlnLeuSerMetGlnMet 1049  
 QY 1598 ACTGTATCAGTTTGTGTTATGCTTTTATGAGCGGCTCCTGTCTTT-----1648  
 Db 1050 SerLeuGlyProLeuValIhISgylMetLeuThSerGlyValAlaValPheMetLeuSer 1069  
 QY 1649 -----AATTGGCATTTATTTAGACCATAGGATTTCTT 1681  
 Db 1070 ThrSerProPheGluPheValIleArgHisPheCysTrpLeuLeuValIleValCys 1089  
 QY 1682 CTAGGAACCTTATCGTCTTTTATGATGACACACCTGTGTTGTTATGATGACCGTA 1741  
 Db 1090 ValGlyAlaCysAsnSerLeuLeuValPhe--ProIleLeuLeuSerMetValGlyPro 1108  
 QY 1742 GAAATATGCTCAAAATATAGCCTTAACCTTAATCTTAACGTAGACATATATAAAATCT 1801  
 Db 1109 Gln-----AlaGluLeuValProLeuGlnHisProAspArgIleSerThr---ProSer 1125  
 QY 1802 CCTTTGGAGCTTTAGTCCCAAGG 1825  
 Db 1126 ProLeuProValArgSerSerLys 1133  
 RESULT 35  
 US-08-460-900C-62  
 ; Sequence 62, Application US/08460900C  
 ; Patent No. 6165747  
 GENERAL INFORMATION:  
 APPLICANT: Ingham, Phillip W.  
 APPLICANT: McMahon, Andrew P.  
 APPLICANT: Tabin, Clifford J.  
 APPLICANT: Bumcrot, David A.  
 APPLICANT: Marti-Gorosita, Elisa  
 TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing  
 TITLE OF INVENTION: Proteins and Uses Related Thereto  
 NUMBER OF SEQUENCES: 62  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FOLEY, HONG & ELIOT LLP  
 STREET: One Post Office Square  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/460,900C  
 FILING DATE: 5-JUNE-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/435,093  
 FILING DATE: 4-MAY-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/356,060  
 FILING DATE: 14-DEC-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/176,427  
 FILING DATE: 30-DEC-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Vincent, Matthew P.  
 REGISTRATION NUMBER: 36,709  
 REFERENCE/DOCKET NUMBER: HMV-006,05  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 832-1000  
 TELEFAX: (617) 832-7000  
 INFORMATION FOR SEQ ID NO: 62:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1299 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear

MOLECULE TYPE: protein  
 US-08-460-900C-62  
 Alignment Scores:  
 Pred. No.: 0.124  
 Score: 93.50  
 Percent Similarity: 42.91%  
 Best Local Similarity: 23.13%  
 Query Match: 2.82%  
 DB: 4  
 Gaps: 14  
 US-09-868-987-1 (1-1864) x US-08-460-900C-62 (1-1299)  
 QY 1082 TTAAGCTATATAGAGATAGACAGCTCTCTCTAAATTAACGATGACGCTGGCGT 1141  
 Db 922 LeuThnAspThSerGlnIleLysThrLeuIleGlyHisIleArgAspLeuSerValLys 941  
 QY 1142 TATTGGGATTTGTTGTCGAAAACAGGCTTGA-----TTTCTTACGGAATCTCT 1192  
 Db 942 TyrGluGlyPheGlyLeuProAsnLysProSerGlyIleProPheIle-----957  
 QY 1193 AAACGAAAGCGAAAATTTTGTCGAAAGTAAAGCAGCAGAACTATCGAAGAAATGCGTAT 1252  
 Db 958 -----PheTrpGluGlnIlyrMetThrLeuArgSerSerLeuAlaMetIle 972  
 QY 1253 CAGCGACCATCGCGCTTTTATGAGCCTTTGGCAATCATC---TTGCTTATGAGATTGG 1309  
 Db 973 LeuAlaCysVal---LeuLeuAlaIleValIleValSerLeuLeuLeuSerVal 991  
 QY 1310 CGCTTGAATGCGCATATGCTTCACTGCGCGTATGCGCTTAAATTCATGACCTTTGGCT 1369  
 Db 992 -----Trp-----AlaAlaValIleValIleLeuSerValIleAla 1003  
 QY 1370 ACCTGTCAGCTCTTTGTTATGACATTTCTTTGAAGAAATTCAAATGATTTGGCA 1429  
 Db 1004 SerLeuAlaGlnIlePheGlyAla-----MetThrLeuLeuGlyIleLysLeuSer 1020  
 QY 1430 GCCATTGCTCTTTAAATGACATGATTTGGGATTCATTAACAAATCTTATGATCTTTT 1489  
 Db 1021 AlileProAlaValIleLeuIleLeuSerValGlyMet-----MetLeuCysPhe 1037  
 QY 1490 GATCGATTTGCGAAGATCGCAGCGAAGCACTGTATACCCCTTATGCATGTTTATGTTAT 1549  
 Db 1038 -----AsnValIleLeuIleSer 1042  
 QY 1550 GATGCCCTTCAAAAAGAGCTTCAGC-----CGCAGGTAATGACACAGCTGACA 1597  
 Db 1043 LeuGlyPheMetThSerValGlyAsnArgGlnArgValGlnLeuSerMetGlnMet 1062  
 QY 1598 ACTGTATCAGTTTGTGTTATGCTTTTATGAGCGGCTCCTGTCTTT-----1648  
 Db 1063 SerLeuGlyProLeuValIhISgylMetLeuThSerGlyValAlaValPheMetLeuSer 1082  
 QY 1649 -----AATTGGCATTTATTTAGACCATAGGATTTCTT 1681  
 Db 1083 ThrSerProPheGluPheValIleArgHisPheCysTrpLeuLeuValIleValCys 1102  
 QY 1682 CTAGGAACCTTATCGTCTTTTATGATGACACACCTGTGTTGTTATGATGACCGTA 1741  
 Db 1103 ValGlyAlaCysAsnSerLeuLeuValPhe--ProIleLeuLeuSerMetValGlyPro 1121  
 QY 1742 GAAATATGCTCAAAATATAGCCTTAACCTTAATCTTAACGTAGACATATATAAAATCT 1801  
 Db 1122 Gln-----AlaGluLeuValProLeuGlnHisProAspArgIleSerThr---ProSer 1138  
 QY 1802 CCTTTGGAGCTTTAGTCCCAAGG 1825  
 Db 1139 ProLeuProValArgSerSerLys 1146  
 RESULT 36  
 US-08-674-509B-48  
 ; Sequence 48, Application US/08674509B  
 ; Patent No. 6261786

GENERAL INFORMATION:  
 APPLICANT: Ingham, Phillip W.  
 APPLICANT: McMahon, Andrew P.  
 APPLICANT: Tabin, Clifford J.  
 APPLICANT: Marigo, Valeria  
 TITLE OF INVENTION: SCREENING ASSAYS FOR HEDGEHOG AGONISTS  
 TITLE OF INVENTION: AND ANTAGONISTS  
 NUMBER OF SEQUENCES: 48  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
 STREET: One Post Office Square  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109-2170

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US 08/460,900  
 FILING DATE: 05-JUN-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Vincent, Matthew P.  
 REGISTRATION NUMBER: 36,709  
 REFERENCE/DOCKET NUMBER: HMV-006.06  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-832-1000  
 TELEFAX: 617-832-7000  
 INFORMATION FOR SEQ ID NO: 48:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1299 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-674-509B-48

Alignment Scores:  
 Pred. No.: 0.124  
 Score: 93.50  
 Percent Similarity: 42.91%  
 Best Local Similarity: 23.13%  
 Query Match: 2.82%  
 DB: 4

US-09-868-987-1 (1-1864) x US-08-674-509B-48 (1-1299)

Qy 1082 TTAAGCTATACATAGCAGATACGAGCCTCTCTCTTAAATTAACGATCATGCTGGGT 1141  
 Db 922 LeuthraspThrSerGlnIleuysThrLeulleGlyHisIleArgAspLeuSerValuys 941  
 Qy 1142 TATTGTGGATGTGTTCAGAAACAGGCGCTAGA-----TTTCTACGGAACCTCT 1192  
 Db 942 TyrGluGlyPheGlyLeuProAsnTyrProSerGlyIleProPheIle----- 957  
 Qy 1193 AACGAACGCAAAATTTTGTCAAGTAGCAGCAAACTATCGAAGAAATGCGTTAT 1252  
 Db 958-----PheTrpGluGlnThrMetThrLeuArgSerSerLeuAlaMetile 972  
 Qy 1253 CAGCGACCATCGGCTTTTAGGAGCTTTGGCAATCATC---TTGCTCTATGTGATTTG 1309  
 Db 973 LeuAlaCysVal---LeuLeuAlaLeuValLeuValSerLeuLeuLeuSerVal 991  
 Qy 1310 CGCTTTGAATGGCAATATGCTTTTCAGTCGCGCTATCGCTTTAAATTCATGACCTTTGGCT 1369  
 Db 992-----Ttp-----AlaAlaValLeuValIleLeuSerValLeuAla 1003  
 Qy 1370 ACCTGTGAGCTCTGTTTATAGCACATTTCTTTTGAAGAAAAATTCAAATAGATTTGCAA 1429

Db 1004 SerLeuAlaGlnIlePheGlyAla-----MetThrLeuLeuGlyIleLysLeuSer 1020  
 Qy 1430 GCCATTGGTCTTTAAATGACTGTATTGGGGTATTTCATTAAACAATACCTTTGATCATTTT 1489  
 Db 1021 AlaIleProAlaValIleLeuIleLeuSerValGlyMet-----MetLeuCysPhe 1037  
 Qy 1490 GATCGTATTCGTGAAGATCGCCAAAGCAACCTGTTTACCCCTATGCATGTTTGTAGTTAAT 1549  
 Db 1038-----AsnValLeuIleSer 1042  
 Qy 1550 GATGCCCTTCAAAAGACGTTTCAGC-----CGCACGGTAATGACACACAGCTACA 1597  
 Db 1043 LeuGlyPheMetThrSerValGlyAsnArgGlnArgValGlnLeuSerMetGlnMet 1062  
 Qy 1598 ACTCATCAGTTTTGTAAATGCTTTTCTTTTATAGCGCGCTCTCTGCTCTTT----- 1648  
 Db 1063 SerLeuGlyProLeuValHisGlyMetLeuThrSerGlyValAlaValPheMetLeuSer 1082  
 Qy 1649-----AATTTGCATTTATTATGACCATAGGAGTCTT 1681  
 Db 1083 ThrSerProPheGluPheValIleArgHisPheCysTrpLeuLeuValValLeuCys 1102  
 Qy 1682 CTAGGAACCTTTATCGTCTCTTTATATTGCACCACTCTGTTGTTGTTTATGTCGCGTAAA 1741  
 Db 1103 ValGlyAlaCysAsnSerLeuLeuValPhe---ProIleLeuLeuSerMetValGlyPro 1121  
 Qy 1742 GAAATCGCTCAAAATAAGTACCGTTAAACTTAATCTAAAGCTGTAGCAATATAAAATCT 1801  
 Db 1122 Glu-----AlaGluLeuValProLeuGluHisProAspArgIleSerThr---ProSer 1138  
 Qy 1802 CTTTGGGACTTTTAGTCCCAAGG 1825  
 Db 1139 ProLeuProValArgSerSerLys 1146  
 RESULT 37  
 US-08-954-698-48  
 ; Sequence 48, Application US/08954698  
 ; Patent No. 6271363  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ingham, Phillip W.  
 ; APPLICANT: McMahon, Andrew P.  
 ; APPLICANT: Tabin, Clifford J.  
 ; TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing  
 ; TITLE OF INVENTION: Proteins and Uses Related Thereto  
 ; NUMBER OF SEQUENCES: 48  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
 ; STREET: One Post Office Square  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02109-2170  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/954,698  
 ; FILING DATE: 20-OCT-1997  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/462,386  
 ; FILING DATE: 05-JUN-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/435,093  
 ; FILING DATE: 04-MAY-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/356,060  
 ; FILING DATE: 14-DEC-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/176,427  
 ; FILING DATE: 30-DEC-1993

ATTORNEY/AGENT INFORMATION:

NAME: Vincent, Matthew H

REGISTRATION NUMBER: 36,103

REFERENCE/DOCUMENT NUMBER: HW-008.10

TELEPHONE: 617-832-1000

TELEPHONE: 017-832-1000  
TELEFAX: 617-832-7000

IDENTIFIER: 01, 002, 1000  
OPERATION FOR SEQ ID NO: 48:

EXPERIENCE CHARACTERISTICS:

LENGTH: 1299 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

954-698-48

**Alignment Scores:**

Pred. No.:	0.124	Length:	129
Score:	93.50	Matches:	62
Percent Similarity:	42.91%	Conservative:	53
Best Local Similarity:	23.13%	Mismatches:	90
Query Match:	2.82%	Indels:	63
DB:	4	Gaps:	14

US-09-868-987-1 (1-1864) x US-08-954-698-48 (1-1299)

Oy	1082	TTAAGCTATACATTAAGCAATGACGAGCGCTCTCTCTAAATTAAGCATGACATGAGCGGCT	1141
Db	922	LeuThrAspHisSerGlnIleuLysThrIleuIleGlyHisIleArgAspLeuSerValLys	941
Oy	1142	TATTTGGGATTTGTGTGCAGAAAACAGGCCCTAGA-----TTTCTCTACGGAAATCTCT	1192
Db	942	TyrGluGluPheGlyLeuProAsnTyrProSerGlyIleProPheIle-----	957
Oy	1193	AAACGAAACCGCAAAATTTTGGTGCAAAAGTATAGCAGAACTATCGAAGAAATGGTTAT	1252
Db	958	-----PheTyrGluGlnTyrMetThrLeuArgSerSerLeuAlaMetIle	972
Oy	1253	CAGCGACCAATCGGGGCTTTTAGAGAGCTTTGGCAATATC---TTGCTATCATGTGATTG	1309
Db	973	LeuAlaCysVal--LeuLeuAlaAlaLeuValLeuValSerLeuLeuLeuSerVal	991
Oy	1310	CGCTTGAATGGCAATATGCTTTTCAGTCCGATAGCGCTTATTAATCATGACCTTTGGCT	1369
Db	992	-----Tyr-----AlaAlaValLeuValIleLeuSerValLeuAla	1003
Oy	1370	ACCTGTGCAGTCTTTGTTATAGCAATTTCTTTGAAGAAATTCAAATATGATTGGCA	1429
Db	1004	SerLeuAlaGlnIlePheGlyAla-----MetThrLeuGluGlyIleLysLeuSer	1020
Oy	1430	GCCATTGTGCTTTAATGACTGTAATTTGGGATTCATTAAACAATCTTGATCATTTTT	1489
Db	1021	AlaIleProAlaValIleLeuIleLeuSer-ValGlyMet-----MetLeuGlySHe	1037
Oy	1490	GATGTATTCGTGAAGATCGCAAGCAACCTGTTACCCATGCGATGGATTTTAACTAT	1549
Db	1038	-----AsnValLeuIleSer	1042
Oy	1550	GATGCCCTTCAAAAGACGTTACG-----CGCACGGTATAGCAACACCTACA	1597
Db	1043	LeuGlyPheMetThrSerValGlyAsnArgGlnArgValGlnLeuSerMetGlnMet	1062
Oy	1598	ACTGTATCAGTTTGTATATGCTTTTGTATATAGCGGGCTCTGTGTTT-----	1648
Db	1063	SerLeuGlyProLeuValHisGlyMetLeuThrSerGlyValAlaValPheMetLeuSer	1082
Oy	1649	-----AATTGGCATTTATATGACCTAGGAGTTCTT	1681
Db	1083	ThrSerProPheGluPheValIleArgHisPheCysTyrPheLeuLeuValIleLys	1102
Oy	1682	CTAGAACCTTATCGCTCTTATATAGCAACACCTGTGTTGTTTATGTCCTGGTAA	1741
Db	1103	ValGlyAlaCysAsnSerLeuValPhe--ProIleLeuSerMetValGlyPro	1121

[illegible]

### RESULT 38

US-09-134-001C-5567

; Sequence 5567, Application US/09134001C

; Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO THE FUNCTION OF THE

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GIC-00/  
CITIZENS ADDICTION NUMBER: 716/00/134 0010

CURRENT FILING DATE 1998-08-13

CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064 964

FRYCK AFFIDAVIT NUMBER: 03 00/004, 204  
PRIOR FILING DATE: 1987-11-08

PRIOR APPLICATION NUMBER: IIS 60/055 779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEO ID NOS: 5674

SEO ID NO 5567

LENGTH: 344

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5567

### Alignment Scores

Pred. No.:	0.0702	Length:	73
Score:	93.00	Matches:	34
Percent Similarity:	37.08%	Conservative:	69
Best Local Similarity:	19.06%	Mismatches:	13
Query Match:	2.81%	Indels:	108
DB:	4	Gaps:	17

US-09-868-987-1 (1-1864) x US-09-134-001C-5567 (1-344)

OY	452	ATTGTCCTGTGCATAGGGGATGGCCCGTGAATGCCAAATTTCTTGATTCGAAAGCAATCCGA	511
Dd	44	LeulleuSerMetSerleuSerleuSerSetrTpaspValGllylysueAapBspValPhe	63
OY	512	GAGGAATTTTATTGTCTCAA--AGTCTTAATAAAATCTGTAGAAAAAGATTTACCAAG	568
Dd	64	AasnInthrIleuPenegInvalArgLeuproArgleuennGuImetleuthrGly	83
OY	569	GCTTT-----GGAGCATTGTTGATCTPAACTGTAGCTACAGATTGGCCTCA	616
Dd	84	AlavalIeuthrLeualaGlyGlnvalrtyGlnlIeValIeuSnsAnpProleuAlaAsp	1030
OY	617	GCACCTCTTTTCTTCCTAGATACAGGGCCATTAAAAGSGTTGCTTGACATTGATTTA	676
Dd	104	SerPhetrHeuGlyLeuAlaSerGlyAlaserleuGlyserGlyIlealaleuPheIeu	1233
OY	677	GGA-----ATTTCCTC-----TCAGTTTAGGCGCTT	706
Dd	124	GlyIeuSerPheIeuTrpPheProIephSerIleIephSerIleuIethIeuIeu	1433
OY	707	TTCATGACATAAATTTTCTCATGCTGTGATGATGAATAGACCACAACATACAGTTGCAT	766
Dd	144	LeuValIeuSerValSerAlaMetIeu-----	152
OY	767	ATGATGAATTAAGTTCGTGGGGATTAACAGATGATTTCTTGAGAGATGACAAAACTTGG	826
Dd	152	-----	152
OY	827	GCTGTTCTCGAAGGTTTTTCTTTAGGTGCGCTTCTCGGGTTTGAGCC---TGG	883
Dd	153	AlaIaySgIyrrProValGImetLeuIIeuthrclyleuIeuPheGlyAlaIeuleu	1727
OY	884	AATTCGTTTTGGGAATGATTTTAAAGAAGGATAGCCCTTTACCTTTAATCCAAAAAG	943

Db 173 AsnAlaLeuTyrLeu-----LeuValLeuIleAsnProLysLys 186  
Qy 944 CATGGCATCAGCGATGTGCTCAAAATGCGTGGCAAAAGTTGTGCATAAACTACAGGAAGCT 1003  
Db 187 -----MetAsnProIleAla-----SerTyrLeuPheGly 196  
Qy 1004 GGTCTTCTCTAGACATTCGGT-----ATTCAACATTTGGATCTTCAGAA 1051  
Db 197 GlyPheAlaSerAlaGluTyrGlnAepValMetIleIleSerLeuIleAlaSerValAla 216  
Qy 1052 AAGATCAAAATCTATTTAGTGATAAAGCTTTAAGCTATATAAG-----CAGATACGA 1105  
Db 217 IleIleValLeuPheLeuMetGlnLysGlyIleLysLeuGlnValGlyIleLeuLys 236  
Qy 1106 GCCTCTCTCTAAATTAACGATCATGAGTCGGCTTATTGTGGGATGTGTGTGAGAAAC 1165  
Db 237 SerGlnSerLeuGlyLeuAsnValGlnGlnValThrTyrIleValLeuIleVal----- 254  
Qy 1166 AGGCTAGATTTCTCTACGGAACTCTAAACGAACGCAAAATTTGGTCAAGGTAAGC 1225  
Db 255 -----AlaSerIleMetThrAlaVal--- 261  
Qy 1256 AGCAAACTATCGAAGAAATGCTTATCAGCGGACCATCGGCTTTTAGGAGCTTTGGCA 1285  
Db 262 -----ValValAlaTyrValGlyValIleGlyPheIleGly 273  
Qy 1286 ATCATPTTGTCTCTATGTGAGTTTGGCTTTGAATGGCAATATGCTTTCAGTCCGCTATGC 1345  
Db 274 MetIleIleProGlnLeuIleArgLysPheTyrTrpAgtTyrGlnIleGly-----Leu 291  
Qy 1346 GCTTAAATCATGACCTTTGGCTACCTGTGCGAGTCTTGTGTATAGCACATTC----- 1399  
Db 292 GlnMetLeuLeuAsnIleIleIleGlyAlaValValMetIleIleAlaAspPheIleGly 311  
Qy 1400 -----TTTTCAAGAAATTCATTAATGATCATTTTGTATGATGCTTTCGTAAGTCTGA 1453  
Db 312 SerThrMetIleGlnProIleGlnIle-----ProValGlyIleValMetAlaLeu 328  
Qy 1454 TTGGGCTATTCAATCAAACTACTTTGATCATTTTGTATGCTGATGCTGTAAGTCTGA 1513  
Db 329 LeuGlyValProValLeuPheTyrIleLeuIle-----LysGln 341  
Qy 1514 GCGAACCTG 1522  
Db 342 ThrAsnIle 344

RESULT 39  
US-08-118-270-73  
; Sequence 73, Application US/08118270  
; Patent No. 5508384  
; GENERAL INFORMATION:  
; APPLICANT: Murphy, Randall B.  
; APPLICANT: Schuster, David I.  
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
; NUMBER OF SEQUENCES: 348  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/118,270  
; FILING DATE: 09-SEP-1993  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/943,236  
; FILING DATE: 10-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Townsend, Kevin G.  
; REGISTRATION NUMBER: 34,033  
; REFERENCE/DOCKET NUMBER: MURPHY=2A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 73:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 334 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-118-270-73

Alignment Scores:  
Pred. No.: 0.0789 Length: 334  
Score: 92.50 Matches: 82  
Percent Similarity: 33.50% Conservative: 54  
Best Local Similarity: 20.20% Mismatches: 115  
Query Match: 2.79% Indels: 155  
DB: 1 Gaps: 22

US-09-868-987-1 (1-1864) x US-08-118-270-73 (1-334)  
Qy 290 ATCTCAGCATGCTGCTGGTGGCAATGCTTATTGTTTGATGAGGGTATATATAGATTT 349  
Db 7 IleLeuIleCysCys-----PheIleIleLeuGluAsnIlePheVal----- 20  
Qy 350 GGAGCGGTCATCGCTGCGGAGCTGTTCTTCTGAATCTTTTGTCTATCTGGGAGCT--- 406  
Db 21 -----LeuLeuThrIleTrrLysThrLys 28  
Qy 407 -----CTACAGTATTGGATGCGCCACTCACCTTTGTCTAGGACTCGCTGGG 451  
Db 29 LysPheHisArgProMetTyrTyrPheIleGlyAsnIleAlaLeuSerAspLeuIleAla 48  
Qy 452 ATTGTTCTGTATGGGATGCGGTAGATGCAAAATGTTCTTGTATTGGAAGAATCCGA 511  
Db 49 -----GlyValAlaTyrThrAlaAsnLeuLeu----- 58  
Qy 512 GAGCAATTTTATTGCTCTCAAAGCTTTAAATAATCTGTAGAAAAGGATATACCAAGSCT 571  
Db 58 ----- 58  
Qy 572 TTGGAGCCATTTTGTATTGATCTTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 631  
Db 59 -----SerGlyAlaThrThrTyrLysLeuThrProAlaGlnTrpPhe 72  
Qy 632 CTAGATACAGGCGCTTAAAGGTTTGTGTTGACATTTGATTGATTGAGNAATTTCTCTTCA 691  
Db 73 LeuArgGluGly-----SerMetPheValAlaLeuSerLeuSerValPheSerLeu 89  
Qy 692 ATGTTTACGCTCTTTTCATGACTAAATTTTCTTCATGCTGTGTGATG-----AATAAG 745  
Db 90 Leu-----AlaIleAlaIleGluArgTyrIleThrMetLeuLysMetLeuHisAsnGly 107  
Qy 746 ACCCAACATACACAGTGTATGATGAATAAGTTCGTGGGGATAAAGCATGATTTCTTGT 805  
Db 108 SerAsnAsnPheArgLeuPheLeu-----Ile 117  
Qy 806 AGAGGATGCAAAAAAATTTGGCTGTTTCTGGAAGTGTTTTCTTTAGTGTGCTGCT 865  
Db 118 SerAlaCys-----TrpValIleSer-----LeuIleLeuGly-----Gly 129  
Qy 866 CTCGGGTTTGGAGCTCGAATTCGTTTGGGATGATTTTAAAGGAGGATGCTGCTTT 925  
Db 130 LeuProIleMetGlyTrpAsnCysIle----- 138

QY 926 ACCTTATATCAAAAGKATGCGATGCGATGCTCAATGCGTGGCAAGTTGTG 985  
 Db 138 ----- 138  
 QY 986 CATTAAGTACAGAGAGCGCTTTCTTTCTAGAGCTTCGATTCAACATTGGATCT 1045  
 Db 139 -----SerAlaLeuSer-----Cys 144  
 QY 1046 TCAGAAAAAGATCAAAATCTATTGATTAAGTAAAGCTTAAGCTAACTAAGACATACGA 1105  
 Db 145 SerThrValLeuProLeuTyrHisIleHisTyrIleLeu--PheCysThrLeuIleVal 163  
 QY 1106 GCCTCTCTCTAAATTAAGATCATGAGCTGGGCTTATGTGGATT-----GTTGTC 1159  
 Db 164 PheThrLeuLeuLeuSerIleValIle--LeuTyrCysArgIleTyrSerLeuVal 182  
 QY 1160 AGAAACAGCGCTTATGTTCTTACGAAACTTAACGAAAGCAAAATTTTGSTCAAG 1219  
 Db 183 ArgThrArgSerArgIleLeu-----ThrPheArgLysAsn----- 194  
 QY 1220 GTAGACGAAACTATCGAAGAAATGCGTTATCAGCGACCATCGGCTTTAGAGCT 1279  
 Db 195 -----IleSerLysAlaSerArgSerSerGluAsnValAlaLeuLeuLysThr 210  
 QY 1280 TTGGCATCATCTTG---CTTATGTAGAGTTGGCTTGAATGGCAATATGCTTTGAGT 1336  
 Db 211 ValIleIleValLeuSerValPheIleAlaCys-----TyrAlaProLeuPheIle 227  
 QY 1337 GCGGTATGCGCTTATGATCATGACCTTTGGCTACCTGTGACCTGCTTATATGACAT 1396  
 Db 228 LeuLeuLeuLeuValGlyCysLysValIleThrCysAspIleLeuPheArgAlaGlu 247  
 QY 1397 TTTCTTTTGAAGAAATTCAAATAGATTTCGACCCATGCTGCTTAATGACTGTATG 1456  
 Db 248 TyrPheLeuVal-----IleAlaValIleAsnSerGlyThrAsnProIle 262  
 QY 1457 GGGTATTCATTAACAT 1474  
 Db 263 IleTyrThrLeuThrAsn 268

RESULT 40  
 PCT-US93-08528-73  
 Sequence 73. Application PC/TUS9308528  
 GENERAL INFORMATION:  
 APPLICANT: New York University  
 TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
 TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
 NUMBER OF SEQUENCES: 348  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: BROWDY AND NEIMARK  
 STREET: 419 Seventh Street, N.W., Suite 300  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20004

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US93/08528  
 FILING DATE: 09-SEP-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/943,236  
 FILING DATE: 10-SEP-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Townsend, Kevin G.  
 REGISTRATION NUMBER: 34,033  
 REFERENCE/DOCKET NUMBER: MURPHY=2 PCT  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-628-5197  
 TELEFAX: 202-737-3528

TELEX: 248633  
 INFORMATION FOR SEQ ID NO: 73:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 334 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 PCT-US93-08528-73

Alignment Scores:  
 Pred. No.: 0.0789 Length: 334  
 Score: 92.50 Matches: 82  
 Percent Similarity: 33.50% Conservative: 54  
 Best Local Similarity: 20.20% Mismatches: 115  
 Query Match: 2.79% Indels: 155  
 DB: 5 Gaps: 22

US-09-868-987-1 (1-1864) x PCT-US93-08528-73 (1-334)

QY 290 ATCTCAGCATGCTGTGGCTTGGCAATGCTTATGTTGATGAGCGTATATTAAGATT 349  
 Db 7 IleLeuIleCysCys-----PheIleIleLeuGluAsnIlePheVal----- 20  
 QY 350 GGAGGCGTATCGCTTCGGAGCGTCTTCTGAATCTTTGCTTATCTGAGGAGCT--- 406  
 Db 21 -----LeuLeuThrIleTyrLys 28  
 QY 407 -----CTACAGATTGATGCGCACCTGCTTACAGGACTCGCTGGG 451  
 Db 29 LysPheHisArgProMetCysTyrPheIleGlyAsnIleAlaLeuSerAspLeuIleAla 48  
 QY 452 ATGTCTTGTCTATGGGAGTGGCGTATGATGCAATGTTCTTATTCGAAGAATCCGA 511  
 Db 49 -----GlyValAlaTyrThrAlaAsnLeuLeu----- 58  
 QY 512 GAGGAATTTTATGTCCTCAAGTCTTAAGAAATCTGTAGAAAAGATATACCAAGCT 571  
 Db 58 ----- 58  
 QY 572 TTTGAGCATTTTGTATCTTAATCTTACTGACTACATGATGCGCTACACATCTTTCTTC 631  
 Db 59 -----SerGlyAlaThrThrTyrLysLeuThrProIleGlnTyrPhe 72  
 QY 632 CTAGATACAGGCGCTTATTAAGGCTTGGCTTGCATGATGTTTGAAGATTCTCTCA 631  
 Db 73 LeuArgGluGly-----SerMetPheValAlaLeuSerLeuSerValPheSerLeu 89  
 QY 692 ATGTTACGCTCTTTTCATGACTTAATTTTCTTCAATGCTGTGATG-----AATAAG 745  
 Db 90 Leu-----AlaIleAlaIleGluArgTyrIleThrMetLeuLysMetLeuHisAsnGly 107  
 QY 746 ACCCAATACACAGTTCATGATGATGATGATGATGCTGTGGGATTAAGCATGATTTCTTG 805  
 Db 108 SerAsnAsnAsnLeuPheLeu-----Ile 117  
 QY 806 AGAGATGCAAAAAAAGCTTGGCGCTGTTCTGAGAGTGTCTTTTCTTCTTCTGCTGCT 865  
 Db 118 SerAlaCys-----TyrValIleSer-----LeuIleGluGly-----Gly 129  
 QY 866 CTGGGCTTGGAGCTTGAATTCCTGTTTGGGAATGATTTTAAAGAGGATATGCTTT 925  
 Db 130 LeuProIleMetGlyTyrPheAsnCysIle----- 138  
 QY 926 ACCTTATATCAAAAGATGATGATGATGATGATGCTCAAAATGCGTGGCAAGTTGTG 985  
 Db 138 ----- 138  
 QY 986 CATTAAGTACAGAGAGCGCTTTCTTTCTAGAGCTTCGATTCAACATTGGATCT 1045  
 Db 139 -----SerAlaLeuSer-----Cys 144  
 QY 1046 TCAGAAAAAGATCAAAATCTATTGATTAAGCTTAAGCTTAAGCTTAAGACATACGA 1105



```
Db 145 SerThrValLeuProLeuTyrHisLysHisTyrIleLeu---PheCysThrLeuIleVal 163
Qy 1106 GCCTCTCTCTAAATACCATCATGAGTGGCGTATTATTTGGGATT-----GTTGTC 1159
Db 164 PheThrLeuLeuLeuSerIleValIle---LeuTyrCysArgIleTyrSerLeuVal 182
Qy 1160 AGAACAGCGCTAGATTCTCTACGGAAACTCTAAACGAAACGCAAAATTTTGGTCAAAG 1219
Db 183 ArgThrArgSerArgArgLeu-----ThrPheArgLysAsn----- 194
Qy 1220 GTAACGACAAATCTCAAGAAATGCGTTATACAGCGGACCATCGGCTTTTAGGAGCT 1279
Db 195 -----IleSerLysAlaSerArgSerSerGluAsnValAlaLeuLeuLysThr 210
Qy 1280 TTGGCAATCATCTTG---CTCTATGTGATTGCGCTTTGTAATGGCAATATGCTTTTCACT 1336
Db 211 ValIleIleValLeuSerValPheIleAlaCys-----TipAlaProLeuPheIle 227
Qy 1337 GCGGTATGCGCTTAAATTCATGACCTTTTGGCTACTGTGTCAGTCTTGTATATACACAT 1396
Db 228 LeuLeuLeuLeuAspValGlyCysLysValLysThrCysAspIleLeuPheArgAlaGlu 247
Qy 1397 TCTCTTTTGAGAAATTCARATAGATTGCAAGCCATTGGTGCTTTAATGACTGTATTG 1456
Db 248 TyrPheLeuVal-----IleAlaValIleAsnSerGlyThrAsnProIle 262
Qy 1457 GGGTATTTCATTAACAAT 1474
Db 263 IleTyrThrLeuThrAsn 268

RESULT 41
US-09-387-699-2
; Sequence 2, Application US/09387699
; Patent No. 6221660
; GENERAL INFORMATION:
; APPLICANT: Bonini, James A.
; APPLICANT: Borowsky, Beth E.
; APPLICANT: Adham, Nika
; TITLE OF INVENTION: DNA Encoding SNORF25 Receptor
; FILE REFERENCE: 56095-A
; CURRENT APPLICATION NUMBER: US/09/387,699
; PRIORITY FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/255,376
; EARLIER FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 2
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-387-699-2

Alignment Scores:
Pred. No.: 0.0903 Length: 335
Score: 92.00 Matches: 65
Percent Similarity: 37.46% Conservative: 59
Best Local Similarity: 19.64% Mismatches: 139
Query Match: 2.78% Indels: 68
DB: 4 Gaps: 15

US-09-868-987-1 (1-1864) x US-09-387-699-2 (1-335)

Qy 434 TTGTGAGGACTCGCTGGGATTCTTCTGTATGGGATGGCGGTAGATGCAATGTTCTT 493
Db 11 LeuAlaValLeuAlaSerLeuIleIleAlaThrAsnThrLeuValAlaValLeu 30
Qy 494 GTATTGCAAGAAATCCGAGAGGAATTTTATTGTCTCAAAAGTCTTAAATAATCTGTAGAA 553
Db 31 LeuLeuIleHisLysAsnAspGlyValSerLeuCysPheThrLeuAsnLeuAlaVal 50
Qy 554 AAAGATATACCAAGCTTTTGGAGCCATTTTGTGATTTCTAACTTGACTACAGTATGGCC 613
```

```
Db 51 -----AspThrLeuIleGlyValAlaIleSerGlyLeuLeuThrAspGlnLeuSer 67
Qy 614 TCAGCAGCTCTTTCTTCCTAGATACAGGGCTATTAAAGGGTTTGTCTTTGACATTGATT 673
Db 68 Ser-----ProSerArgProThrGlnLysThrLeuCysSerLeuArg 81
Qy 674 TTAGAATTTTCTCTCAATGTTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTG 733
Db 82 MetAlaPheValThrSerSerAlaAlaSerValLeuThr-----ValMet 97
Qy 734 TGGATGAATAGACCCACATACAGCTTGTCATATGATGAATAAGTTGCTGGGATAAAG 793
Db 98 LeuIleThrPheAspArgTyrLeuAlaIleLysGlnProPheArgTyrLeuLysIleMet 117
Qy 794 CATGATTTCTTTGAGAGGATGC-----AAAAAACTTTGGGCTGTTCTTGGAAAGTGT 847
Db 118 SerGlyPheValAlaGlyAlaCysIleAlaGlyLeuTrpLeuValSer-----Tyr 134
Qy 848 CTTTATGTTGCTGCTCTCGGGTTTGGAGCCCTGGAAATCCGTTTGGGAATGGAATTTT 907
Db 135 LeuIleGlyPheLeuProLeuGlyIleProMetPheGlnGlnThrAla-----Tyr 151
Qy 908 AAAGGAGGATGCTCTT-----ACCTTTAATCCAAAAGAGCAT----- 946
Db 152 LysGlyGlnCysSerPhePheAlaValPheHisPro-----HisPheValLeuThrLeu 169
Qy 947 -----GGCATCAGC 955
Db 170 SerCysValGlyPhePheProAlaMetLeuLeuPheValPhePheTyrCysAspMetLeu 189
Qy 956 GATGTTGCTCAATGCGTGGCAAGTTGTGCATAAACTACAGGAAGCTGCTCTTTCTTCT 1015
Db 190 LysIleAlaSerMetHisSerGlnGlnIleArgLysMetGluHisAlaGlyAlaMetAla 209
Qy 1016 AGAGACTTCCGTATT-----CAACATTTGGATCTTCAGAAAAAGATCAAAATC 1063
Db 210 GlyGlyTyrArgSerProArgThrProSerAspPheLysAlaLeuArgThrValSerVal 229
Qy 1064 TATTTAGTGATAAAGCTTTAAGCTATATACAGAGATCAGGAGCTCTCTCTAAAATTA 1123
Db 230 LeuIleGlySerPheAlaLeuSerTrpThrProPheLeuIleThrGlyIleValGlnVal 249
Qy 1124 ACGATCATGAGCTGGCGTTATTGTGGGATTGTTGTGAGAAACAGGCTAGATTCTCTAC 1183
Db 250 AlaCysGlnGluCysHisLeuTyrLeuValLeu-----GluArgTyrLeuTrp 265
Qy 1184 -----GGAAACTCTAAAGAAACGCA-----AAATTTGGTCAAAG--- 1219
Db 266 LeuLeuGlyValGlyAsnSerLeuLeuAsnProLeuIleTyrAlaTyrTrpGlnLysGlu 285
Qy 1220 GTAAGCAGCAAACTATCGAAGAAATGCGTTATCAGGCGACCATCGGCTTTTAGGAGCT 1279
Db 286 ValArgLeuGlnLeu-----TyrHisMetAlaLeuGlyValLysLysVal 300
Qy 1280 TTGGCAATCATCTTGTCTCTATGTGAGTTTGGCG 1312
Db 301 LeuThrSerPheLeuLeuPheLeuSerAlaArg 311

RESULT 42
US-09-641-259B-2
; Sequence 2, Application US/09641259B
; Patent No. 6468756
; GENERAL INFORMATION:
; APPLICANT: Bonini, James A
; APPLICANT: Borowsky, Beth E
; APPLICANT: Adham, Nika
; APPLICANT: Boyle, No. 64687561
; APPLICANT: Thompson, Thelma O.
; TITLE OF INVENTION: DNA Encoding SNORF25 Receptor
; FILE REFERENCE: 1995/56095-B/JPW/ADM
; CURRENT APPLICATION NUMBER: US/09/641,259B
; CURRENT FILING DATE: 2002-03-12
```

PRIOR APPLICATION NUMBER: PCT/US00/04413  
 PRIOR FILING DATE: 2000-02-22  
 PRIOR APPLICATION NUMBER: US 09/387,699  
 PRIOR FILING DATE: 1999-08-13  
 PRIOR APPLICATION NUMBER: US 09/255,376  
 PRIOR FILING DATE: 1999-02-22  
 NUMBER OF SEQ ID NOS: 35  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 2  
 LENGTH: 335  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-641-259B-2

Alignment Scores:  
 Pred. No.: 0.0903 Length: 335  
 Score: 92.00 Matches: 65  
 Percent Similarity: 37.46% Conservative: 59  
 Best Local Similarity: 19.64% Mismatches: 139  
 Query Match: 2.78% Indels: 68  
 Gaps: 15

US-09-868-987-1 (1-1864) x US-09-641-259B-2 (1-335)

QY 434 TTGTCAGAGACTCGCGATTGTTCTGCTATGGGAGTGGCCCTGATGCAATGTTCTT 493  
 DB 11 LeuAlaValLeuAlaSerLeuLeuLeuAlaThrAsnThrLeuValAlaValAlaValLeu 30  
 QY 494 GTATTCAGAAAGATCCGAGAGAGATTTTATGTCGCAAGCTTAAATAATCTGTGAA 553  
 DB 31 LeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 50  
 QY 554 AAAGATATACCAAGCTTTTGAGCCATTTTATGATTCATCTGACTACAGTATGGCC 613  
 DB 51 -----AspThrLeuLeuLeuLeuValAlaLeuSerGlyLeuLeuThrAspGlnLeuSer 67  
 QY 614 TCAGCACTTCTTTCTTCCTGATACAGAGGCTATTAAGGTTGCTTACATGATTT 673  
 DB 68 Ser-----ProSerArgProThrGlnGlySerThrLeuCySerLeuArg 81  
 QY 674 TTGGAATTTCTCTCAATGTTTACGGCTCTTTCATGACTAAATTTTCTTCATGCTG 733  
 DB 82 MetAlaPheValInrSerSerAlaAlaAlaSerValLeuInr-----ValMet 97  
 QY 734 TGGATGATTAAGCAACATACACAGTTCATGATGATTAAGTTCGTGGGATTAAG 793  
 DB 98 LeuLeuThrPheAspArgTyrLeuAlaAlaLeuGlnProPheArgTyrLeuLeuIleMet 117  
 QY 794 CATGATTTCTGAGAGATGC-----AAAAAATCTTGGGCTGTTCTGGAAGTCTTTT 847  
 DB 118 SerGlyPheValAlaGlyAlaCysIleAlaGlyLeuTyrPheValSer-----Tyr 134  
 QY 848 CTTTAAAGTTCGCTGCTCGGCTGGTTGGAGCCGATTCGTTTGGGAATGATTTT 907  
 DB 135 LeuLeuGlyPheLeuProLeuGlyIleProMetPheGlnInrAla-----Tyr 151  
 QY 908 AAAGAGGATGATGCTTT-----ACCTTATATCAAAAAGCAT----- 946  
 DB 152 LysGlyGlnCysSerPhePheAlaValPheHisPro-----HisPheValLeuThrLeu 169  
 QY 947 -----GGCATCAGC 955  
 DB 170 SerCysValGlyPhePhePheAlaMetLeuLeuPheValPhePheTyrCysAspMetLeu 189  
 QY 956 GATGTTCTCAAAAGCGG33AAAGTTGTCATAACACAGAAAGCTGCTTCTTCTT 1015  
 DB 190 LysLeuAlaSerMetHisSerGlnGlnInrLeuArgLysMetGlnHisAlaGlyAlaMetAla 209  
 QY 1016 AGAGATTCGCTATTT-----CAAAATTTGATCTTCAAGAAAGATCAAAATC 1063  
 DB 210 GlyGlyTyrArgSerProArgThrProSerAspPheLysAlaLeuArgThrValSerVal 229  
 QY 1064 TATTTAGTATAAGCTTTAAGCTATATCAAGAGATACGAGGCTCTCTCTTAAATTA 1123

DB 230 LeuIleGlySerPheAlaLeuSerTrpThrProPheLeuLeuThrGlyIleValGlnVal 249  
 QY 1124 ACGATCATGAGTGGCGTTATTTGGGATTTGTCAGAAAGCGCTGATTTCTCTAC 1183  
 DB 250 AlaCysGlnGluCysHisLeuTyrLeuValLeu-----GlnArgTyrLeuTyrP 265  
 QY 1184 -----GGAACTCTAAAGCAAGCGCA-----AAATTTGGTCAAG--- 1219  
 DB 266 LeuLeuGlyValGlyAsnSerLeuLeuAsnProLeuIleTyrAlaTyrTrpGlnLysGln 285  
 QY 1220 GTAGAGCAAACTATCGAAGAAATGCTTATCAGCGCAGCATCGGCTTTAGAGAGCT 1279  
 DB 286 ValArgLeuGlnLeu-----TyrHisMetAlaLeuGlyValLysVal 300  
 QY 1280 TTGCAATCATCTTGCTCTATGATGATTTGGCC 1312  
 DB 301 LeuThrSerPheLeuLeuPheLeuSerAlaArg 311

RESULT 43

US-08-118-270-64

Sequence 64, Application US/08118270  
 Patent No. 5508384

GENERAL INFORMATION:

APPLICANT: Murphy, Randall B.

TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN

TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF

NUMBER OF SEQUENCES: 348

CORRESPONDENCE ADDRESS:

ADDRESSER: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/118,270

FILING DATE: 09-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/943,236

FILING DATE: 10-SEP-1992

ATTORNEY/AGENT INFORMATION:

NAME: Townsend, Kevin G.

REGISTRATION NUMBER: 34,033

REFERENCE/DOCKET NUMBER: MURPHY=2A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

TELEX: 248633

INFORMATION FOR SEQ ID NO: 64:

SEQUENCE CHARACTERISTICS:

LENGTH: 269 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-118-270-64

Alignment Scores:  
 Pred. No.: 0.105 Length: 269  
 Score: 91.00 Matches: 44  
 Percent Similarity: 37.82% Conservative: 29  
 Best Local Similarity: 22.80% Mismatches: 70  
 Query Match: 2.75% Indels: 50  
 Gaps: 8

US-09-868-987-1 (1-1864) x US-08-118-270-64 (1-269)





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 17, 2002, 11:55:41 ; Search time 20 Seconds  
(without alignments)  
7731.188 Million cell updates/sec

Title: US-09-868-987-1  
Perfect score: 3312  
Sequence: 1 atggacttcgcattgtc.....atgacaattcagataatgc 1864

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-Q=/cgn2\_1/USFTO\_spool/US09868987/runat\_16122002\_083006\_23528/app\_query.fasta\_1.2055  
-DB=SwissProt\_40 -QFMT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=45 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USR=US09868987 @CGN 1 1 13 @runat\_16122002\_083006\_23528 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAR -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WAEN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	443	13.4	472	1	SECD_SYNY3 Q55610 synechocyst
2	431	13.0	503	1	SECD_HELPY O26074 helicobacte
3	429	13.0	526	1	SECD_HELPJ O2zj66 helicobacte
4	427	12.9	501	1	SECD_AQUAE O67102 aquifex aeo
5	426	12.9	583	1	SECD_TREPA O83425 treponema p
6	418.5	12.6	554	1	SECD_RHOCA O33517 rhodobacter
7	408.5	12.3	616	1	SECD_HABIN P44591 haemophilus
8	407.5	12.3	518	1	SECD_BRUAB O2zg86 brucella ab
9	398.5	12.0	518	1	SECD_RICPR O9zcw8 rickettsia
10	364	11.0	323	1	SECF_ECOLI P19674 escherichia
11	361	10.9	333	1	SECF_RHOCA O33518 rhodobacter
12	360	10.9	324	1	SECF_RHOSH O33568 rhodobacter
13	359.5	10.9	615	1	SECD_ECOLI P19673 escherichia
14	355.5	10.7	615	1	SECD_SALCH O9zff8 salmonella
15	351	10.6	323	1	SECF_HELPY O2zj65 helicobacte
16	350	10.6	323	1	SECF_HELPY O26073 helicobacte
17	342	10.3	311	1	SECF_RICPR O9ze34 rickettsia
18	342	10.3	325	1	SECF_HABIN P44590 haemophilus

19	333.5	10.1	586	1	SECD_BORBU
20	313	9.5	288	1	SECF_AQUAE
21	304	9.2	571	1	SECD_MYCLE
22	303	9.1	573	1	SECD_MYCTU
23	291.5	8.8	299	1	SECF_BORBU
24	279.5	8.4	442	1	SECF_MYCTU
25	276.5	8.3	570	1	SECD_STRCO
26	269.5	8.1	420	1	SECF_TREPA
27	257	7.8	315	1	SECF_SYNY3
28	256	7.7	373	1	SECF_STRCO
29	252	7.6	471	1	SECF_MYCLE
30	183.5	5.5	396	1	Y111_MENJA
31	156.5	4.7	403	1	Y843_METHA
32	147	4.4	574	1	MUP1_YEAST
33	144	4.3	1034	1	ACRF_ECOLI
34	136.5	4.1	1032	1	Y895_HABIN
35	135	4.1	847	1	MLLB_STRCO
36	132	4.0	282	1	YC53_MENJA
37	129	3.9	1040	1	YEGN_ECOLI
38	128	3.9	944	1	MLL3_MYCTU
39	125.5	3.8	955	1	MLL3_MYCLE
40	125.5	3.8	1037	1	YHIV_ECOLI
41	124.5	3.8	1014	1	MLLB_MYCLE
42	123.5	3.7	1025	1	YEGO_ECOLI
43	123.5	3.7	1046	1	MEXB_PSEAE
44	121.5	3.7	966	1	MLLB_MYCTU
45	118.5	3.6	1037	1	ACRD_ECOLI

# ALIGNMENTS

RESULT 1	SECD_SYNY3	STANDARD;	PRT;	472 AA.
AC	Q55610;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Protein-export membrane protein sec.			
GN	SECD OR SLR0774.			
OS	Synechocystis sp. (strain PCC 6803)			
OC	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.			
OX	NCBI_TaxID=1148;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96127529; PubMed=8590279;			
RA	Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,			
RA	Sugiura M., Tabata S.;			
RT	"Sequence analysis of the genome of the unicellular cyanobacterium			
RT	Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb			
RT	region from map positions 64* to 92* of the genome.";			
RL	DNA Res. 2:153-166(1995)			
CC	-1- FUNCTION: INVOLVED IN PROTEIN EXPORT.			
CC	-1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS			
CC	WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY			
CC	(BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).			
CC	-1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.			

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib/announce/>  
or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
-----

EMBL; D64000; BAA10118.1;  
DR InterPro; IPR001035; Acrlflin res.  
DR InterPro; IPR003335; Secd\_Secf.  
DR Pfam; PF02335; Secd\_Secf; 1.  
DR PRINTS; PR00702; ACRIFLAVINRP.  
DR TIGRFAMs; TIGR00916; 2A0604s01; 1.

DR TIGRAME; TIGR01129; sec2, 1.  
 KW Protein transport; Translocation; Transmembrane; Membrane;  
 FT Complete proteome.  
 FT TRANSMEM 7 27 POTENTIAL.  
 FT TRANSMEM 298 318 POTENTIAL.  
 FT TRANSMEM 335 355 POTENTIAL.  
 FT TRANSMEM 411 431 POTENTIAL.  
 FT TRANSMEM 432 452 POTENTIAL.  
 SQ SEQUENCE 472 AA; 50468 MW; D246FEF2DCB263CA CRC64;

Alignment Scores:  
 Pred. No.: 4,99e-28 Length: 472  
 Score: 443.00 Matches: 89  
 Percent Similarity: 62.39% Conservative: 57  
 Best Local Similarity: 38.03% Mismatches: 82  
 Query Match: 13.38% Indels: 6  
 Gaps: 1

US-09-868-987-1 (1-1864) x SEC2\_SYN3 (1-472)

QY 50 CAATATTCTGCAACCGTGAATGCGCTGATGAGCGTTATATGCTAGC 109  
 Db 213 GlnAlaValAlaGlyThrGlyArgSerLeuGlyValPheLeuAspAsnAspLeuIleSer 232  
 QY 110 AGCCCTATTTAAACGTCCTTGAATAAT-----CATGCCAGTCTC 151  
 Db 233 AlaProValAlaGlyValGluPheAlaAsnThrGlyIleThrGlyAlaAlaValIle 252  
 QY 152 TCAGCAAAATTTACCCACCGCTGAATGAGCAAACTCCGCTGATTTAAATCTGAGCG 211  
 Db 253 ThrGlyAsnThrIleAspThrAlaAsnAspLeuAlaGlnLeuArgGlyGlySer 272  
 QY 212 ATGCTTTTGTCCGAGCTTCAGTGAAGACAGCATCTTCTGATCTTTGGAAAAA 271  
 Db 273 LeuProPheProValGluValAlaGlnAsnArgThrValGlyAlaThrLeuGlyGlnGlu 292  
 QY 272 CAATGTACACAGGCAATTTCTGACATGCTGCTGCGCTTGGCAAGCTTATTTGATG 331  
 Db 293 SerIleArgArgSerLeuValAlaGlyPheValGlyLeuValLeuValPheMet 312  
 QY 332 AGCGTATATATATGATTTGAGGCGTCATCGGAGCGTCTTCTGATCTTTG 391  
 Db 313 AlaValIleThrArgLeuProGlyIleValAlaAspIleSerLeuMetIleThrAlaVal 332  
 QY 392 CTATCTGGGAGCTCTACAGTATTTGGAGCGCACTCACCTTGTCAAGACTCGCTGG 451  
 Db 333 LeuThrLeuAlaAlaPheAlaLeuValGlyValThrLeuThrLeuProGlyIleAlaGly 352  
 QY 452 ATTCTTTCTGATGCGGATGCGCTGATGCAAAATGTTCTGATTCGAAGAATCCGA 511  
 Db 353 PheIleLeuSerIleGlyMetAlaValAspAlaAsnValLeuIlePheGluArgThrArg 372  
 QY 512 GAGGAATTTTATTTGCTCAAGACTTAAATAATCTGTGAAGAAAGATACCAAGGCT 571  
 Db 373 GluIleLeuAlaGlyAlaGlyAsnThrLeuThrArgSerValGluAlaGlyPhePheArgAla 392  
 QY 572 TTGGAGCCATTTTGAATCTTAATCTGACTACAGTATTTGGCTGAGCATCTTTCTCTC 631  
 Db 393 PheSerSerIleLeuAspSerAsnValThrThrLeuIleAlaCysAlaAlaLeuPheThr 412  
 QY 632 CTAGATCAGGGCCTATTTAAGGCTTGGCTTTGACATTTAGATTTAGAAATTTCTCTCA 691  
 Db 413 PheGlySerGlyLeuValIleGlyPheAlaLeuThrLeuAlaIleGlyValMetValSer 432  
 QY 692 ATGTTACGGCTCTTTTCATGATCAATAATTTTCTTCATGCTG 733  
 Db 433 LeuPheThrAlaLeuThrCysSerArgThrLeuLeuVal 446

RESULT 2  
 SEC2\_HELPY STANDARD; PRT; 503 AA.  
 ID 026074;  
 DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein-export membrane protein sec2.  
 GN SEC2 OR HP1550.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE=97394467; PubMed=9252185;  
 RA Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G.,  
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.R., Peterson S.,  
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 RA Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,  
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,  
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Frazer C.M.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the gastric pathogen Helicobacter  
 pylori.";  
 RL Nature 388:539-547(1997).  
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).  
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS  
 CC WHICH COMPRISE SEC2, SEC3, SEC4, SEC5, SEC6, SEC7, SEC8 AND SEC9  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SEC2/SEC3 FAMILY. SEC2 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AB000652; AAD08588.1; -  
 DR TIGR; HP1550; -  
 DR InterPro; IPR001036; Acrlflavin res.  
 DR InterPro; IPR003335; Sec2 Sec3.  
 DR Pfam; PF02355; Sec2 Sec3\_1.  
 DR PRINTS; PR00702; ACRIFLAVINRP.  
 DR TIGRAMES; TIGR00916; ZAD064801.1.  
 DR TIGRAMEs; TIGR01129; sec2, 1.  
 KW Protein transport; Translocation; Transmembrane; Membrane;  
 KW Complete proteome.  
 FT TRANSMEM 199 219 POTENTIAL.  
 FT TRANSMEM 334 354 POTENTIAL.  
 FT TRANSMEM 357 377 POTENTIAL.  
 FT TRANSMEM 383 403 POTENTIAL.  
 FT TRANSMEM 456 476 POTENTIAL.  
 SQ SEQUENCE 503 AA; 54247 MW; 8541C291CA3117086 CRC64;

Alignment Scores:  
 Pred. No.: 4,66e-27 Length: 503  
 Score: 431.00 Matches: 99  
 Percent Similarity: 60.83% Conservative: 47  
 Best Local Similarity: 41.25% Mismatches: 92  
 Query Match: 13.01% Indels: 2  
 Gaps: 2

US-09-868-987-1 (1-1864) x SEC2\_HELPY (1-503)

QY 47 GGAATATTCT---GGAACCGTGAATGCGCTGATGAGCGTTATATG 103  
 Db 252 GlyAspPheSerGlyAlaAsnValGlyLysArgMetAlaIleValLeuAspAsnLysVal 271  
 QY 104 GTGACAGCCCTATTTAAACGTCCTTGAATAATTTTCTTCATGCTGAGGAAA 160  
 Db 272 TySerIleProValIleArgGluArgIleGlyGlySerGlyGlnIleSerGlyAsn 291

QY 161 TTTACCACCGTGAAGTACGCAAACTCGCTCAGATTTAAATCTGAGCGATCTCTTT 220  
 Db 292 PheSerValAlaGlnAlaSerAspLeuAlaValleAlaLeuArgSerGlyAlaMetSerAla 311  
 QY 221 GTTCCCGAGGTTCTCAGTGAGAGACGATCTCTCTGATCTTTGGGAAAAACAATGTACA 280  
 Db 312 ProIleGlnValLeuGluLeuArgIleGlyProSerLeuGlyLysAspSerValLys 331  
 QY 281 CAAGGCATTATCTCAGCATGCTGTGGCTGGCAATGCTTATTTGATGAGCGTATAT 340  
 Db 332 ThrSerIleAlaLeuValGlyGlyPheIleLeuValMetGlyPheMetValLeuTyr 351  
 QY 341 TATAGATTGGAGCGCTCATCGCTCGGAGCTGTTCTCTGAACTCTTCTTATCTG 400  
 Db 352 TyrSerMetAlaGlyValIleAlaCysLeuAlaLeuValValAsnLeuPheLeuVal 371  
 QY 401 GCAGCTCTACAGTATTGGATGGCCCACTACCTTGTCCAGGACTCGCTGGGATGTTCTT 460  
 Db 372 AlaValMetAlaIlePheGlyAlaThrLeuThrLeuProGlyMetAlaGlyIleValLeu 391  
 QY 461 GCTATGGGATGGCGGTAGTACCAATGTTCTCTGATTTGCAAGAAATCCGAGAGCAATTT 520  
 Db 392 ThrValGlyIleAlaValAlaAspAlaAsnIleIleAsnGluArgIleArgGluValLeu 411  
 QY 521 TTATTGCTCAAGCTTTAAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCC 580  
 Db 412 ArgGluAsnGluGlyIleAlaLysAlaIleHisLeuGlyTyrIleAsnAlaSerArgAla 431  
 QY 581 ATTTTGTATCTCACTGACTACAGTATGTCCTGCTCAGCAGCTCTTCTTCTCTAGATACA 640  
 Db 432 IlePheAspSerAsnIleThrSerLeuIleAlaSerValLeuLeuTyrAlaTyrGlyThr 451  
 QY 641 GGGCTATTAAAGGTTGCTTTCACATGATGTTTAAAGAAATTTCTCTCAATGTTTACG 700  
 Db 452 GlyAlaIleLysGlyPheAlaLeuThrThrGlyIleGlyIleLeuAlaSerIleIleThr 471  
 QY 701 GCTCTTTTCATGACTAAATTTTCTCATGCTGTGGATGAATAGACCAACATACACAG 760  
 Db 472 AlaIleValGlyThrGlnGlyIleTyrGlnAlaLeuLeuProLysLeuThrGlnThrLys 491

RESULT 3  
 ID SECD\_HELPJ STANDARD; PRT; 526 AA.  
 AC Q9ZJ66;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein-export membrane protein.  
 GN SECD OR JHP1449.  
 OS Helicobacter pylori J99 (Campylobacter pylori J99).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 OX NCBI\_TaxID=85963;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99120557; PubMed=9923682;  
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,  
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,  
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,  
 RA Trust T.J.;  
 RT "Genomic sequence comparison of two unrelated isolates of the human  
 RT gastric pathogen Helicobacter pylori.";  
 RL Nature 397:176-180(1999)  
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).  
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS  
 CC WHICH COMPRISE SECA, SECD, SECE, SECF, SECG AND SECY  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; AB001567; AD07024.1; -.  
 DR InterPro; IPR001036; Acrlflvin\_res.  
 DR InterPro; IPR003335; SecD SecF.  
 DR Pfam; PF02355; SecD SecF; 1.  
 DR PRINTS; PR00702; ACRIFLAVINRP.  
 DR TIGRFAMs; TIGR00916; 2A0604s01; 1.  
 DR TIGRFAMs; TIGR01129; secD; 1.  
 KW Protein transport; Translocation; Transmembrane; Membrane;  
 KW Complete proteome.  
 FT TRANSMEM 8 28 POTENTIAL.  
 FT TRANSMEM 356 376 POTENTIAL.  
 FT TRANSMEM 379 399 POTENTIAL.  
 FT TRANSMEM 453 473 POTENTIAL.  
 FT TRANSMEM 478 498 POTENTIAL.  
 SQ SEQUENCE 526 AA; 56796 MW; 5B4C9E905794782A CRC64;

Alignment Scores:  
 Pred. No.: 6,798-27 Length: 526  
 Score: 429.00 Matches: 97  
 Percent Similarity: 60.83% Conservative: 49  
 Best Local Similarity: 40.42% Mismatches: 92  
 Query Match: 12.95% Indels: 2  
 DB: 1 Gaps: 2

US-09-868-987-1 (1-1864) x SECD\_HELPJ (1-526)

QY 47 GGACAATATTCT--GCAAAACCGTGAATGCGTATGCTGTAGTATTGACGGTTATATG 103  
 Db 274 GlyAspPheSerGlyAlaAsnValGlyLysArgMetAlaIleValLeuAspAsnLysVal 293  
 QY 104 GTGAGCAGCCCTATTTTAAACGTCCTCATG---AAAAATCATGCCAGTGTCTCAGGAAA 160  
 Db 294 TyrSerAlaProValIleArgGluArgIleGlyGlySerGlyGlnIleSerGlyAsn 313  
 QY 161 TTTACCCACCGTGAAGTACGCAAACTCGCTCAGATTTAAATCTGAGCGATCTCTTT 220  
 Db 314 PheSerValAlaGlnAlaSerAspLeuAlaIleAlaLeuArgSerGlyAlaMetAsnAla 333  
 QY 221 GTTCCCGAGGTTCTCAGTGAGAGACGATCTCTCTGATCTTGGGAAAAACAATGTACA 280  
 Db 334 ProIleGlnValLeuGluLysArgIleValGlyProSerLeuGlyLysAspSerIleLys 353  
 QY 281 CAAGGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTATTTGATGAGCGTATAT 340  
 Db 354 ThrSerIleIleAlaLeuValGlyGlyPheIleLeuValMetGlyPheMetAlaLeuTyr 373  
 QY 341 TATAGATTGGAGCGCTCATCGCTCGGAGCTGTTCTCTGAACTCTTCTTATCTG 400  
 Db 374 TyrSerMetAlaGlyValIleAlaCysMetAlaLeuValValAsnLeuPheLeuVal 393  
 QY 401 GCAGCTCTACAGTATTGGATGGCCCACTACCTTGTCCAGGACTCGCTGGGATGTTCTT 460  
 Db 394 AlaValMetAlaIlePheGlyAlaThrLeuThrLeuProGlyMetAlaGlyIleValLeu 413  
 QY 461 GCTATGGGATGGCGGTAGATGCAAAATGTTCTTCTTTCGAAAAGAAATCCGAGAGCAATTT 520  
 Db 414 ThrValGlyIleAlaValAlaAspAlaAsnIleIleIleAsnGluArgIleArgGluValLeu 433  
 QY 521 TTATTGCTCAAGCTTTAAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCC 580  
 Db 434 ArgGluGlyGluGlyValValLysAlaIleHisLeuGlyTyrIleAsnAlaSerArgAla 453  
 QY 581 ATTTTGTATCTCACTGACTACAGTATGTCCTGCTCAGCAGCTCTTCTTCTCTAGATACA 640  
 Db 454 IlePheAspSerAsnIleThrSerLeuIleAlaSerValLeuLeuTyrAlaTyrGlyThr 473

```

QY 641 GGGCATTAAAGGGTTCTTGACATTGATTTAGGAATTTCTTCAATGTTTACG 700
DB 474 GYAlaIleIysGlyPheAlaLeuThrThrGlyIleGlyIleLeuAlaSerIleIleThr 493
QY 701 GCTCTTTTCAGACTAAATTTTCTTCAGCTGTCGTAATTAAGCCCAACATACAGAC 760
DB 494 AlaIleIleGlyThrGlnGlyIleTyroIleAlaLeuLeuProIysLeuAlaGlnThrIlys 513

RESULT 4
SECD_AQUAE STANDARD; PRT; 501 AA.
ID SECD_AQUAE
AC 06102;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-export membrane protein secd.
GN SECD OR AQ 973.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VFS;
RX MEDLINE=98196666; PubMed=9537320;
RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aubay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS
CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000716; AAC0706G.1; -.
DR InterPro; IPR003335; SECD_SecF.
DR Pfam; PF02355; SECD_SecF; 1.
DR TIGRFAMs; TIGR00916; ZM0604S01; 1.
DR TIGRFAMs; TIGR01129; secd; 1.
KW Protein transport; Translocation; Transmembrane; Membrane;
KM Complete proteome.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 339 359 POTENTIAL.
FT TRANSMEM 371 391 POTENTIAL.
FT TRANSMEM 394 414 POTENTIAL.
FT TRANSMEM 470 490 POTENTIAL.
SQ SEQUENCE 501 AA; 55459 MM; E129FC97445D6C97 CRC64;

Alignment Scores:
Pred. No: 9.78e-27 Length: 501
Score: 427.00 Matches: 82
Percent Similarity: 63.84% Conservative: 61
Best Local Similarity: 36.61% Mismatches: 81
Query Match: 12.89% Indels: 0
Gaps: 0

US-09-868-987-1 (1-1864) x SECD_AQUAE (1-501)
QY 62 AACGTGATGGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121
DB 269 AsnIleGlyIysArgLeuAlaIleValLeuAspArgIysValSerAlaProValIle 288

```

```

QY 122 AACCTCCATTGAAAAATCATGCCAGTCTCAGGAAATTTACCAACCGTAGAGC 181
DB 289 ArgSerIysIleSerAspArgGlyGlnIleThrGlyAsnIleThrAlaGlnAlaArg 308
QY 182 AAATCGCCCTCAGATTAATAATCGAGCGATGCTTTGTCGCCAGTCTCAGTAA 241
DB 309 AspLeuAlaLeuIleLeuAlaGlnThrGlySerLeuProSerProLeuIysPheLeuGlnIu 328
QY 242 GAGACGATCTCTTGTGATCTGGAAAAAACAATGACACAAAGCATTAATTCAGATGC 301
DB 329 LysIleValGlyProSerLeuGlyIysAspAlaIleGlnGlnGlyIleValAlaGlyIle 348
QY 302 TGTGCTTGCGCAATGCTTAATTTGATGATGACGCTAATTAATTAATTTGAGCGGTATC 361
DB 349 LeuAlaIleLeuLeuAlaValAlaLeuIleAlaArgTyrosThrAlaGlyIleThr 368
QY 362 GCTTCGGAGCTGTTCTTGAATCTTTGATCTTATCTGAGCGAGCTACAGATTTGAT 421
DB 369 AlaAsnIleSerIlePheLeuAsnValLeuPheLeuLeuAlaSerMetAlaPheLeuGly 388
QY 422 GCGCCACTCAGCTTGTGAGACTCGCTGGATGTTCTGTAATGGAGTCCGATGAT 481
DB 389 AlaThrLeuThrLeuProGlyIleAlaGlyIleIleLeuAsnMetGlyIleAlaValAsp 408
QY 482 GCAAAATGTTCTTGATTCGAAAGAAATCCGAGAGAAATTTATTTGCTCAAAAGCTTAA 541
DB 409 SerAsnValIleLeuIlePheGlyArgValIysGlnIleuArgLeuGlyAsnThrValSer 428
QY 542 AAATCTGTAGAAAAAGATATACCAAGCTTTTGAGCCATTTTGTATCTTAATCTGACT 601
DB 429 LysAlaIleGlyLeuGlyPheIysArgThrLeuSerAlaValIlePhePThrIleValThr 448
QY 602 ACAGTATGCGCTCAGACCTTTTCTTCAGATACAGAGGCTATTAAGGTTTGT 661
DB 449 LeuLeuValAlaSerValIleLeuPheGlnPheGlySerGlyProValIysGlyPheAla 468
QY 662 TTGACATTGATTTAGGAATTTCTCTCAATGTTTACGCGCTTTTCATGACTAAATT 721
DB 469 ThrThrLeuAlaLeuGlyThrIleAlaSerPheIleSerAsnValTyrosAlaVal 488
QY 722 TTCTTCATGCTG 733
DB 489 PheLeuAspLeu 492

RESULT 5
SECD_TREPA STANDARD; PRT; 583 AA.
ID SECD_TREPA
AC 083425;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-export membrane protein secd.
GN SECD OR TP0410.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RL Science 281:375-388(1998).
CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT.
CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS

```



CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY  
(BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----

DR EMBL; AE001219; AAC65398.1; -;  
DR TIGR; TP0410; -;  
DR InterPro; IPR003335; SecD SecF.  
DR Pfam; PF02355; SecD SecF; 1.  
DR TIGRFAMS; TIGR00916; 2A0604s01; 1.  
DR TIGRFAMS; TIGR01129; secD; 1.  
KW Protein transport; Translocation; Transmembrane; Membrane;  
KW Complete proteome.  
FT TRANSMEM 7 27 POTENTIAL.  
FT TRANSMEM 419 439 POTENTIAL.  
FT TRANSMEM 453 473 POTENTIAL.  
FT TRANSMEM 511 531 POTENTIAL.  
FT TRANSMEM 538 558 POTENTIAL.  
SQ SEQUENCE 583 AA; 63860 MW; 9A787E64C4CFE1A1 CRC64;

## Alignment Scores:

Pred. No.: 1.2e-26 Length: 583  
Score: 426.00 Matches: 92  
Percent Similarity: 64.25% Conservative: 50  
Best Local Similarity: 41.63% Mismatches: 77  
Query Match: 12.86% Indels: 2  
DB: 1 Gaps: 2

US-09-868-987-1 (1-1864) x SECD\_TREPA (1-583)

QY 68 GGATGGCGTGTAGTGTAGTGTAGTGTATGTGTGTCAGCAGCCCTATTAAAGCTC 127  
DB 343 GlyArgGluAlaIleValSerAspGlyIleArgSerAlaProAlaIleArgGlu 362  
QY 128 CCATTG---AAAATCATGCCAGTGTCTCAGGAAATTTACCCCGGTGAAGTGAGCAAA 184  
DB 363 ProIleThrAlaGlySerGlySerIleSerGly---PheSerAlaGluGluAlaGlnAsn 381  
QY 185 CTCGCCCTCAGATTAAATCTGGAGCGATGCTTTTGTTCGCCGAGTCTCAGTGAAGAG 244  
DB 382 LeuIysThrAlaLeuArgSerAlaTrpLeuAsnValAlaLeuGluIleGluAsnGlnGln 401  
QY 245 ACGATCTCTCTGATCTTGGGAAAACAAATGTACACAGGCATTATCTCAGCATGCTGT 304  
DB 402 ValValGlyAlaSerMetGlyGluGluSerIleArgGlnGlyThrArgAlaLeuValTrp 421  
QY 305 GGCTTGGCAATGCTTATTTGTTGATGAGCGTATATATATAGATTGGAGCGCTCATCGCT 364  
DB 422 GlyLeuCysAlaValLeuLeuPheMetLeuValTrpTyrGlnGluAlaGlyValAsnAla 441  
QY 365 TCGGAGCTGTCTTCTGAATCTTTGCTTATCTGGGAGCTCTACAGTATTTGGATCGG 424  
DB 442 CysValAlaGlnLeuLeuAsnLeuTyrIleMetPheGlyValLeuSerAlaPheAsnLeu 461  
QY 425 CCACCTCCTGTGTGAGGACTCGCTGGGATTTCTTGTGATGGGATGGCGGTAGATGCA 484  
DB 462 ThrLeuThrLeuSerSerIleAlaGlyMetIleLeuThrIleGlyMetAlaValAspAla 481  
QY 485 AATGTTCTGTATTGCGAAGAAATCCGAGGAAATTTTATTTGTTCTCAAGTCTTAAAAA 544  
DB 482 AsnValValValPheGluArgIleArgGluGluLeuAlaLeuGlyLysSerArgGlyAla 501  
QY 545 TCTGTAGAAAAGGATATACCAAGCTTTTGGAGCCATTTTGTGATCTCAATCTGACTACA 604  
DB 502 AlaValCysSerGlyPheGluArgAlaPheTrpAlaIleMetAspSerAsnValThrThr 521

QY 605 GTATTGGCTCAGCAGCTTCTTTCTTCCTAGATACAGGGCCTATTAAAGGTTTGTCTTG 664  
DB 522 PheIleAlaAlaLeuPheLeuSerValLeuGlyThrGlyProIleLysGlyPheAlaTyr 541  
QY 665 ACATTGATTTAGGAATTTCTCTCAATGTTTACGCGCTCTTTTCATGACTAAATTTTC 724  
DB 542 SerLeuAlaIleGlyValValSerValPheThrAlaLeuPheValSerArgLeuMet 561  
QY 725 TTC 727  
DB 562 Phe 562

## RESULT 6

SECD\_RHOCA STANDARD; PRT; 554 AA.  
ID SECD\_RHOCA  
AC O33517;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Protein-export membrane protein secD.  
GN SECD.  
OS Rhodobacter capsulatus (Rhodopseudomonas capsulata).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
OC Rhodobacter.  
OX NCBI\_TaxID=1061;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SB1003 / St Louis;  
RX MEDLINE=97318920; PubMed=9175857;  
RA Goldman B.S., Beckman D.L., Ball A., Monika E.M., Gabbert K.K.,  
RA Kranz R.G.;  
RT "Molecular and immunological analysis of an ABC transporter complex  
RT required for cytochrome c biogenesis.";  
RL J. Mol. Biol. 268:724-738(1997).  
CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).  
CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS  
CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY  
(BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----

DR EMBL; U69979; AAB62801.1; -;  
DR InterPro; IPR003335; SecD SecF.  
DR Pfam; PF02355; SecD SecF; 1.  
DR TIGRFAMS; TIGR00916; 2A0604s01; 1.  
DR TIGRFAMS; TIGR01129; secD; 1.  
KW Protein transport; Translocation; Transmembrane; Membrane.  
FT TRANSMEM 10 30 POTENTIAL.  
FT TRANSMEM 392 412 POTENTIAL.  
FT TRANSMEM 435 455 POTENTIAL.  
FT TRANSMEM 491 511 POTENTIAL.  
FT TRANSMEM 516 536 POTENTIAL.  
SQ SEQUENCE 554 AA; 58943 MW; DF2CBEBBA9F69EDF CRC64;

## Alignment Scores:

Pred. No.: 4.79e-26 Length: 554  
Score: 418.50 Matches: 86  
Percent Similarity: 61.92% Conservative: 62  
Best Local Similarity: 35.98% Mismatches: 88  
Query Match: 12.64% Indels: 3  
DB: 1 Gaps: 3

US-09-868-987-1 (1-1864) x SECD\_RHOCA (1-554)

QY 29 ATCAGCGGCACT---GCTAATGACAAATATCTGCAACCGT---GGATGGCGTATGGCT 82  
 Db 305 ValSerGlyAlaArgAlaPheGlyAspTyrThrAlaGlyHisIleGlyGluProPheAla 324  
 QY 83 GTATGATATGAGCTTATATGCTGACGACCGCTTATTTAAACGTCCTCAATGAAAATCAT 142  
 Db 325 IleValIleuAspGlyValValIleSerAlaProThrIleGlnAlaHisIleAlaGly 344  
 QY 143 GCCAGT---GTCTCAGGAAATTTACCCACCGTGAAGTGAACAACTCGCTCAGATT 199  
 Db 345 SerGlyIleIleThrGlyArgPheSerIleGluGlnAlaThrAspLeuAlaIleuLeu 364  
 QY 200 AAATCTGAGGAGGATCTTTTGTCCGAGGTCTCAGTGAAGAGAGATCTTTCAT 259  
 Db 365 ArgAlaGlyAlaLeuProAlaGlyMetThrPheLeuGluGlnArgThrIleGlyProGlu 384  
 QY 260 CTTGGGAAAAAACAATGACACAAAGCATATCTCAGCATCTGCTGGCTGGCAATGCTT 319  
 Db 385 LeuGlyAlaAspSerValIleValAlaGlyMetValAlaSerValIleGlyPheValAlaVal 404  
 QY 320 ATGTGTTGATGAGCGTATATTATATGATTTGAGGCGTCATCGCTTGGAGCTGTCTT 379  
 Db 405 ValAlaTyrMetIleAlaSerTyrGlyLeuPheGlyPhePheSerSerValAlaLeuPhe 424  
 QY 380 CTGAATCTTTTGTCTTATCT36GCGACTTACAGTATTTGGATGGCCCACTACCTTGCA 439  
 Db 425 IleAsnIleAlaPheIlePheAlaValMetGlyAlaIleGlyGlyThrMetThrLeuPro 444  
 QY 440 GGACGCGCGGAGATGTTCTTGTCTCTATGCGATGCGCTAGATGCAATGTTCTGTATTC 499  
 Db 445 GlyIleAlaGlyIleValLeuThrIleGlyThrSerValAspAlaAsnValLeuIleYr 464  
 QY 500 GAAAGATCCAGAGAGAAATTTTATTTGCTCAAGCTTTAAATAATCTGTAAGAAAAGA 559  
 Db 465 GluArgMetArgGluGluIleArgSerGlySerProValArgAlaIleGluLeuGly 484  
 QY 560 TATACCAAGCTTTTGGAGCCATTTTATCTTAACTGATCTAGCATTTGGCCCTCAGA 619  
 Db 485 PheAspTyrAlaMetSerAlaIleIleAspAlaAsnValThrSerPheLeuSerSerAla 504  
 QY 620 CTTCTTTTCTTCTAGATACAGAGGCTTATTAAGGGTTGCTTGACATTTGATTAAGA 679  
 Db 505 IleuPheValIleGlyAlaGlyProValArgGlyPheAlaValThrThrMetIleGly 524  
 QY 680 ATTTTCTCTCATGTTTACGCTCTTTTCTGACTAAATTTTCTTCATGCTGTG 736  
 Db 525 IleAlaIleSerIlePheThrAlaIleThrValValArgLeuMetIleValIleThr 543

RESULT 7  
 SECD\_HAEIN STANDARD; PRT; 616 AA.  
 ID SECD\_HAEIN STANDARD; PRT; 616 AA.  
 AC P44591;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-export membrane protein secD.  
 GN SECD OR H10240.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=127;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KM20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom B., Cotton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

RA Venter J.C.;  
 RT "whole-genome random sequencing and assembly of Haemophilus influenzae  
 RL Science 269:496-512 (1995).  
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).  
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS  
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY, SECD FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U32710; AAC21908.1; -.  
 CC TIGR: H10240; -.  
 CC InterPro: IPR003335; SecD SecF.  
 CC Pfam: PF02355; SecD SecF.1.  
 CC TIGRPFAM: TIGR00916; 2A0604501; 1.  
 CC DR TIGRPFAM: TIGR01129; secD; 1.  
 CC KW Protein transport; Translocation; Transmembrane; Inner membrane;  
 CC Complete proteome.  
 CC FT TRANSMEM 11 31 POTENTIAL.  
 CC FT TRANSMEM 453 473 POTENTIAL.  
 CC FT TRANSMEM 475 495 POTENTIAL.  
 CC FT TRANSMEM 497 517 POTENTIAL.  
 CC FT TRANSMEM 562 582 POTENTIAL.  
 CC FT TRANSMEM 585 605 POTENTIAL.  
 CC SQ SEQUENCE 616 AA; 66986 MW; 7DE1E4B08505F8F CRC64;  
 Alignment Scores:  
 Pred. No.: 3.1e-25 Length: 616  
 Score: 408.50 Matches: 87  
 Percent Similarity: 65.85% Conservative: 48  
 Best Local Similarity: 42.44% Mismatches: 59  
 Query Match: 12.33% Indels: 11  
 DB: 1 Gaps: 2  
 US-09-868-987-1 (1-1864) x SECD\_HAEIN (1-616)  
 QY 137 AATCATCCGAGTGTCTCAGGAAATTT-----ACC 166  
 Db 395 AsnValAlaThrIleGlnGlyArgPheGlySerAsnPheGlnIleThrGlyValAspSer 414  
 QY 167 CACCGTAAGTGAACAAACTGCGCTCAGATTAAATCTGAGGAGATGCTTTGTGCC 226  
 Db 415 IleAlaGluAlaHisAsnLeuSerThrLeuLeuysSerGlyAlaLeuIleAlaProIle 434  
 QY 227 GAGGTTCTCAGTGAAGAGACAGATCTCTTGATCTTGGGAAAAACAATGACACAGGC 286  
 Db 435 GlnIleValGluGlnArgThrIleGlyProSerLeuGlyAlaGlnAsnValGluGlnGly 454  
 QY 287 ATATCTCAGAGATCTGTGCTTGGCAATGCTTATTGTTGATGAGCGATATTATTAAGA 346  
 Db 455 IleAsnAlaSerLeuThrProGlyLeuValAlaValIleAlaPheMetLeuPheTyrTyrIys 474  
 QY 347 TTGGAGGCGTCATCGCTTGGGAGCTGTCTTCTGAACTTTTGTCTTATCTGGGAGCT 406  
 Db 475 MetPheGlyValIleAlaSerPheAlaLeuValIleAsnIleValLeuLeuValGlyLeu 494  
 QY 407 CTACAGTATTTG---GATGGCCACTACCTGTGACAGATCTCGTGGAGATTTCTTGCT 463  
 Db 495 MetSerIleuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeuThr 514  
 QY 464 ATGGGAGTGGCGCGAGATGCAAAATGTTCTGTGATTCGAAAGAAATCCGAGAGGAAATTTTA 523  
 Db 515 LeuGlyMetSerValAspAlaAsnValLeuIlePheGlnArgIleGlyGluGluIleArg 534

QY 524 TTGCTCAAGTCTTAAATACTGTAGAAAAGGATATACCAAGGCTTTTGAGGCAATT 583  
 Db 535 AsnGlyArgSerIleGlnAlaIleAsnGluGlyTyrAsnGlyAlaPheThrSerIle 554  
 QY 584 TTTGATCTTAACCTGACTACAGTATTGGCTCAGCACTTCTTTCTTCCTAGATACAGGG 643  
 Db 555 PheAspAlaAsnLeuThrIleLeuThraIleLeuThraIleLeuThraIleGlyThrgly 574  
 QY 644 CTTATTAAAGGGTTGCTTGGACATGATTAGGAATTTTCTCTCAATGTTTACGGCT 703  
 Db 575 ProIleGlnGlyPheAlaIleThrLeuSerLeuGlyValAlaIleSerMetPheThrAla 594  
 QY 704 CTTTTCATGACTAAA 718  
 Db 595 IleThrGlyThrArg 599

RESULT 8  
 ID SECD\_BRUAB STANDARD; PRT; 518 AA.  
 AC Q9ZG86;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Protein-export membrane protein secD (Fragment).  
 GN SECD.  
 OS Brucella abortus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_TaxID=235;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2308;  
 RX MEDLINE=99043892; PubMed=9826342;  
 RA Vemulapalli R., Duncan A.J., Boyle S.M., Sritanganathan N., Toth T.E.,  
 RA Schurig G.G.;  
 RT "Cloning and sequencing of yajC and secD homologs of Brucella abortus  
 RT and demonstration of immune responses to yajC in mice vaccinated with  
 RT B. abortus RB51.";  
 RL Infect. Immun. 66:5684-5691(1998).  
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).  
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS  
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF085217; AAC83691.1; -  
 CC InterPro; IPR001036; Acrlflvin\_res.  
 CC InterPro; IPR003335; SecD SecF.  
 CC Pfam; PF02335; SecD SecF.1  
 CC PRINTS; PR00702; ACRIFLAVINRP.  
 CC TIGRFAMs; TIGR00916; 2A0604801; 1.  
 CC TIGRFAMs; TIGR01129; secD; 1.  
 CC Protein transport; Translocation; Transmembrane; Membrane.  
 CC TRANSMEM 9 29 POTENTIAL.  
 CC TRANSMEM 277 297 POTENTIAL.  
 CC TRANSMEM 301 321 POTENTIAL.  
 CC TRANSMEM 392 412 POTENTIAL.  
 CC TRANSMEM 460 480 POTENTIAL.  
 CC NON\_TER 518 518  
 CC SEQUENCE 518 AA; 55685 MW; 35250B66429A7B3B CRC64;

Alignment Scores:

Pred. No.: 3.66e-25 Length: 518

RESULT 9

Score: 407.50 Matches: 96  
 Percent Similarity: 56.58% Conservative: 76  
 Best Local Similarity: 31.58% Mismatches: 107  
 Query Match: 12.30% Indels: 25  
 DB: 1 Gaps: 7

US-09-868-987-1 (1-1864) x SECD\_BRUAB (1-518)

QY 38 ACTGCTAATGGACAATATTCTGCAAAACCGTGGATGCGGTATGGCTGTAGTATGACGGT 97  
 Db 202 ThrAlaGlnGlyAsnGluAsnSer-----PheAlaIleValValAspAsn 216  
 QY 98 TATATGGTCAGCAGCCCTATTTTAAACGTCCTCCATTG---AAAAATCATGCCAGTGCTCA 154  
 Db 217 GlnValValSerAlaProThrValSerGlyProLeuAspThrSerGluLeuGlnIleGlu 236  
 QY 155 GGGAAATTTTACCACCGTGAAGTGAGCAAACTCGCTCAGATTAAATACTGGAGCGATG 214  
 Db 237 GlyAlaPheAspLeuGlnAlaAlaAsnAsnMetAlaValValLeuArgSerGlyAlaLeu 256  
 QY 215 TCTTTTGTCCCGAGGTTCTCAGTGAAGAGACGATCTTCTCTGATCTTGGGAAAAACAA 274  
 Db 257 ProGlnAlaValThrValLeuGluArgThrIleAlaSerAlaLeuGlyGluAspTyr 276  
 QY 275 TGTACACAGGCATTTATCTCAGCATCTGTGGCTTGGCAATGCTTATGTTTGTATGAGC 334  
 Db 277 AlaSerAlaAlaValLeuAlaAlaLeuLeuAlaAlaValValGlyLeuPheMetVal 296  
 QY 335 GTATATTATAGATTTCGAGGCGTCATCGTTCGGGAGCTGTCTCTGTAATCTTTGCTT 394  
 Db 297 LeuSerTyrGlyIleLeuGlyValIleAlaLeuValAlaLeuValValAlaIleValIle 316  
 QY 395 ATCTGGGAGCTCTCAGATGATTTCGAGTGGCCACTCACCTTGTGAGGATCTCGTGGGATT 454  
 Db 317 LeuThraAlaValLeuSerLeuIleGlyAlaSerIleSerLeuAlaSerIleAlaGlyLeu 336  
 QY 455 GTTCTTGTCTATGGGATGGCGGTAGATGCMAATGTTCTTGTATTGCAAGAATCCGAGAG 514  
 Db 337 ValLeuThrIleGlyLeuAlaValAspAlaHisIleLeuIleTyrGluArgValArgGlu 356  
 QY 515 GAATTTTATTGTCTCAAGTCTTAAATAATCTGTAGAAAAGGATATACCAAGGCTTTT 574  
 Db 357 AspArgArgGlyTyrSerValValGlnAlaMetGluSerGlyPheTyrArgAlaLeu 376  
 QY 575 GGAGCCATTTTGTATCTAATCTGACTACAGTATTGGCTCAGCACTCTTTTCTTCCTTA 634  
 Db 377 SerThrIleValAspAlaAsnLeuThrLeuIleAlaAlaLeuValLeuPheLeuLeu 396  
 QY 635 GATACAGGCGCTATTAAAGGGTTGCTTGCATGATGATTTAGGAATTTTCTCTCAATG 694  
 Db 397 GlySerGlyThrValHisGlyPheAlaLeuThrValAlaIleGlyIleGlyThrThrLeu 416  
 QY 695 TTTACGGCTCTTTTCATGACTAAATTTTCTTCTCATGCTGTGGATGAAT-----AAGACC 748  
 Db 417 PheThrThrLeuThrPheThrArgLeuLeuIleAlaGlnTrpValArgThrAlaLysPro 436  
 QY 749 CAACATACACAGTTGCTATATGATGAATAAGTTCGTGGGATAAAGCAT---GATTTCTTG 805  
 Db 437 LysGluValProLysArgLeuLysLeuValProThrValThrHisIleProPheMet 456  
 QY 806 AGAGGATGCAAAAAAATTTGGGCTGTTCTGGAAGTGTCTTTTCTTTAGGTGTC----- 859  
 Db 457 Arg-----LeuGlnPheValThrLeuGlyIleSerValLeuAlaCysAlaIle 472  
 QY 860 -----GTTGCTCTCGGGTTTGGAGCTCGGAATTCGGTTTGGGAATGGAT 904  
 Db 473 ValValAlaLeuPheValAlaAsnIleGlyPheAsn-----TyrGlyIleAsp 487  
 QY 905 TTTTAAAGGAGGG 916  
 Db 488 PheArgGlyGly 491

```

SECD_RICPR
ID SECD_RICPR STANDARD: PRT; 518 AA.
AC Q92CW8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-export membrane protein secd.
OS SECD OR RP586.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichenitz-Ponten T., Almark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "the genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
CC -! FUNCTION: INVOLVED IN PROTEIN EXPORT.
CC -! SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH
CC (BY SIMILARITY).
CC -! SUBCELLULAR LOCATION: Integral membrane protein (by similarity).
CC -! SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ235272; CAA15031.1; -
DR InterPro; IPR003335; SecD_SecF.
DR Pfam; PF02355; SecD_SecF.1.
DR TIGRFAMs; TIGR00916; 2A0604s01.1.
DR TIGRFAMs; TIGR01129; secd.1.
KW Protein transport; Translocation; Transmembrane; Membrane;
KW Complete proteome.
FT TRANSMEM 9
FT TRANSMEM 29
FT TRANSMEM 356
FT TRANSMEM 376
FT TRANSMEM 377
FT TRANSMEM 397
FT TRANSMEM 406
FT TRANSMEM 426
FT TRANSMEM 464
FT TRANSMEM 484
FT TRANSMEM 486
FT TRANSMEM 506
SQ SEQUENCE 518 AA; 56712 MW; 86D6FC2A5B5DE2AA CRC64;

```

```

Alignment Scores:
Pred. No.: 1,956-24 Length: 518
Score: 398.50 Matches: 86
Percent Similarity: 58.8% Conservative: 54
Best Local Similarity: 36.1% Mismatches: 97
Query Match: 12.0% Indels: 1
DB: Gaps: 1

```

US-09-868-987-1 (1-1864) x SECD\_RICPR (1-518)

```

QY 62 AACCGTGATGGCGTAGTGTAGTGAAGTTATATGTCAGACCCCTATTTTA 121
DB 281 AenThnglyLysArgLeuAlaValLeuAspAsnLysLeuLeuSerAlaProIleIle 300
QY 122 AAC---GTCCTATGAAAAATATATGCGCATGCTCTACGGAAATTTTACCAACCGTGAAGTG 178
DB 301 AenGlyAlaIleIleGlyCylAsnGlyIleIleThnglyAsnPhenThrIleGluSerAla 320
QY 179 AGCAACGCTGCTGATTAATTAATCTGGAGGAGTGTCTTTGTTCCGAGGTTCTCACT 238
DB 321 AsnGluLeuAlaLeuLeuLeuAlaValGlySerLeuProThrProLeuLysIleIleGlu 340

```

```

QY 229 GAGAGAGCATCTCTTGATCTTGGAAAAAACAATGTACACAGCATTATCTACGA 298
DB 341 GlnTrsSerIleGlyProAsnLeuGlyAlaAspSerIleGluSerGlyLysAlaIle 360
QY 229 TGCTGTGGCTTGCAATGCTATTATTTTGGATGAGCGTATATTAATTTGAGGCGTC 358
DB 361 LeuIleGlyPheValAlaValCysIlePheMetIleLeuSerTyrglyValIleGlyLeu 380
QY 359 ATGCGTTGGAGGCTGTTCTTCTGAATCTTTTGGTATTCTGGGAGCTTACAGTATTG 418
DB 381 PheAlaAsnIleAlaLeuIleLeuAlaLeuLeuAlaLeuLeuSerLeuPhe 400
QY 419 GATGGCCACTTCACCTTGTACGACGCTGCGGATTTGTTCTGATAGGGATGCCGTA 478
DB 401 GlnAlaThrLeuThrLeuProGlyIleAlaIleIleLeuThrIleGlyMetAlaVal 420
QY 479 GATGCAATGTTCTTGTATTCGAAGAATCCGAGAGGAAATTTTATGTCACAAAGTCTT 538
DB 421 AspAlaAsnValLeuIleTyrglyIleGlySerGluGluLeuHisLysGlyValSerAsn 440
QY 539 AAAAAATCTGTAGAAAAAGATATACCAAGCCTTTTGGAGCCATTTTGATTTCACTTG 598
DB 441 LeuTyralaIleArgThnglyPheGluSerAlaPheAlaThrIleIleAspSerAsnIle 460
QY 599 ACTACAGTATGGCGCTCAGACCTTCTTCTTCTAGATATACGGCCATTTAAAGGTTT 658
DB 461 ThrThrLeuValAlaAlaPheAlaLeuTyrllePheGlyValGlyAlaIleLysGlyPhe 480
QY 659 GCTTGAATGATGATTTAGAAATTTTCTCAATGTTTACGGCTTTTCATGATGATGATA 718
DB 481 AlaValAlaLeuThrIleGlyIleIleSerSerMetPheSerAlaIleIleThrIlys 500
QY 719 TTTTCTTCATGCTGTGATGAATTAAGCCACATACACAGTTGATATGATG 772
DB 501 LeuLeuIleAspValTrrpValLysTyrllePheLysProLysLysLeuGlyLeuLeu 518

```

```

RESULT 10
ID SECF_ECOLI STANDARD: PRT; 323 AA.
AC P19674; P71113;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-export membrane protein secf.
GN SECF OR B0409 OR Z0508 OR ECS0460.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=9106014; PubMed=2170107;
RA Gardel C., Johnson K., Jacq A., Beckwith J.;
RT "The secd locus of E. coli codes for two membrane proteins required
RT for protein export.";
RL EMBO J. 9:3209-3216(1990).
RN [2]
RP ERRATUM.
RX MEDLINE=9106534; PubMed=2249673;
RA Gardel C., Johnson K., Jacq A., Beckwith J.;
RL EMBO J. 9:4205-4206(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.U.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";

```





CC -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS  
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH  
 CC (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U83136; AAB61904.1; -  
 DR InterPro; IPR003335; SecD\_SecE.  
 DR Pfam; PF02335; SecD\_SecF; 1.  
 DR TIGRFAMs; TIGR00916; 2A0604s01; 1.  
 DR TIGRFAMs; TIGR00966; 3A0501s07; 1.  
 KW Protein transport; Translocation; Transmembrane; Membrane.  
 FT TRANSMEM 16 36 POTENTIAL.  
 FT TRANSMEM 146 166 POTENTIAL.  
 FT TRANSMEM 174 194 POTENTIAL.  
 FT TRANSMEM 266 286 POTENTIAL.  
 SQ SEQUENCE 324 AA; 35088 MW; F4CD14CSB814370E CRC64;  
 Alignment Scores:  
 Pred. No.: 2,34e-21 Length: 324  
 Score: 360.00 Matches: 100  
 Percent Similarity: 49.06% Conservative: 56  
 Best Local Similarity: 31.45% Mismatches: 100  
 Query Match: 10.87% Indels: 62  
 DB: 1 Gaps: 10  
 US-09-868-987-1 (1-1864) x SECF\_RHOSH (1-324)  
 Qy 839 AGTGTTTCTTTAGTGGCTGCTCTCGGTTTGGAGCTGAATTCGGT----- 892  
 Db 27 SerValPheLeuMet-----AlaSerLeuValAlaTrpLeuThrLeuGlyLeu 43  
 Qy 893 ---TTGGCAATGATTTAAAGGGGATGCTTTTACCTTTAATCCAAAGAGCATGCG 949  
 Db 44 AsnPheGlyIleAspPheArgGly-----ThrThrIleArgThrGluSerThrGln 61  
 Qy 950 ATCAGCGATGTTGCTCAAAATGCGTGGCAAGTTGTGCATAACTACAGGAAGCTGGTCTT 1009  
 Db 62 AlaValAspValAlaAlaTyArg-----AlaAlaLeu 72  
 Qy 1010 TCTTCTAGAGAC-----TTCCGTATT 1030  
 Db 73 GluGlyGlnAspLeuGlyAspIleSerIleThrGluValPheAspProGlyPheArgAla 92  
 Qy 1031 CAACATTTGGATCTTCAGAAAGATCAAAATCTATTTTAGTGATAAAGCTTTAGCTAT 1090  
 Db 93 AspGlnHisValAla-----MetValArgIleGlyAlaGlnAspAlaThrGlnSerIle 110  
 Qy 1091 ACTAAG-----CAGATCAGCGCTCTCTCTATAAAATTAACGATCATGAGCTGG 1138  
 Db 111 ThrProGluGlnIleGlyGlnValGluGluAlaLeuIleThrValAsp----- 126  
 Qy 1139 CGTTATTCTGGATTGTTGTGAGAAACAGGCGCTAGATTCTCTACGGAACCTCTAAACGA 1198  
 Db 127 -----ProSerIleThrPheProSerValGlu--- 135  
 Qy 1199 AACGCAAAATTTTGTCTCAAGGTAAAGCAGCAAACTATCGAAGAAATCGGTTATCAGCG 1258  
 Db 136 -----SerValGlyProIleValSerGlyGluLeuIleArgSerAla 149  
 Qy 1259 ACCATCGGGCTTTTAGGCAATCATCTTGGCTCTATGTGAGTTTGGCTTTGCA 1318  
 Db 150 IleIleuAlaValAlaAlaCysAlaGlyIleAlaValIleTrpLeuArgGlu 169  
 Qy 1319 TGGCAATATGCTTTTCAGTGCCGTATGCGCTTTAATTCATGACCTTTTGGCTACCTGTGCA 1378

Db 170 TrpGlnPheAlaLeuGlySerValAlaAlaLeuIleHisaspValLeuValThrIleGly 189  
 Qy 1379 GTCTTGTATTATAGCACATTTCTTTTGAAGAAATTCAAATAGATTTCGAAGCCCATGGT 1438  
 Db 190 Val-----PheAlaLeuPheGlnIleIlePheAspLeuThrThrValAla 204  
 Qy 1439 GCTTTAATGACTGATTTGGGTATTCATTAACAATACTTTTGATCATTTTGTATGCGATT 1498  
 Db 205 AlaLeuLeuThrValLeuGlyIleSerIleAsnAspThrValValPheAspArgLeu 224  
 Qy 1499 CGTCAAGATCGCAAGCAACCTGTTTACCCCTATGCATGTTTGTAGTTAATGATGCCCTT 1558  
 Db 225 ArgGluAsnLeuValIleValIleValIleValIleValIleValIleValIleVal 244  
 Qy 1559 CAAAGACGTTTCAGCCGCGTAAATGACACACAGCTACACTCTATCAGTTTGTATG 1618  
 Db 245 AsnGluThrLeuSerArgThrIleMetThrLeuMetThrThrLeuIleAlaLeuValSer 264  
 Qy 1619 CTTTGTATTATAGCGGCTCTCTCTCTTTAATTTTGCATTTATATGACCATAGGATT 1678  
 Db 265 LeuLeuValPheGlyGlyAspValIleArgGlyPheValPheAlaIleThrPheGlyVal 284  
 Qy 1679 CTTCTAGAACTTTATCGTCTCTCTTATATTCACCACTCTGTTGTTGTTTATG 1732  
 Db 285 ValIleGlyThrIleSerSerValIleMetAlaLeuValIleValLeuValLeu 302  
 RESULT 13  
 SECD\_ECOLI STANDARD; PRT; 615 AA.  
 AC P19673; P77531; P72348;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein-export membrane protein secd.  
 GN SECD OR B0408 OR Z0507 OR ECS0459.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562, 83334;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=K12;  
 RX MEDLINE=9106014; PubMed=2170107;  
 RA Gardel C., Johnson K., Jacq A., Beckwith J.;  
 RT "The secd locus of E.coli codes for two membrane proteins required  
 for protein export.";  
 RL EMBO J. 9:3209-3216(1990).  
 RN [2]  
 RP ERRATUM.  
 RX MEDLINE=91065354; PubMed=2245673;  
 RA Gardel C., Johnson K., Jacq A., Beckwith J.;  
 RL EMBO J. 9:4205-4206(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;







RA Kuhn F.C., Gies A.J., Smeltzer M., Crupper S.S., Sobieski R.J.;  
RL "Identification of secD gene from Salmonella choleraesuis."  
RT Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBUNIT: INVOLVED IN PROTEIN EXPORT.  
CC -!- FUNCTION: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS  
CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY  
CC (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
CC -!- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF106111; AAC83934.1; -;  
DR InterPro; IPR003335; SecD\_SecF.  
DR Pfam; PF02355; SecD\_SecF; 1.  
DR TIGRFAMS; TIGR00916; 2A0604s01; 1.  
DR TIGRFAMS; TIGR01129; secD; 1.  
KW Protein transport; Translocation; Transmembrane; Inner membrane.  
FT TRANSMEM 10 30 POTENTIAL.  
FT TRANSMEM 452 472 POTENTIAL.  
FT TRANSMEM 504 524 POTENTIAL.  
FT TRANSMEM 564 584 POTENTIAL.  
SQ SEQUENCE 615 AA; 66632 MW; 81370B11EFC983F9 CRC64;  
  
Alignment Scores:  
Pred. No.: 5,81e-21 Length: 615  
Score: 355.50 Matches: 80  
Percent Similarity: 61.50% Conservative: 59  
Best Local Similarity: 35.40% Mismatches: 84  
Query Match: 10.73% Indels: 3  
DB: 1 Gaps: 2  
  
US-09-868-987-1 (1-1864) x SECD\_SALCH (1-615)  
QY 44 AATGGCAATATTCGCAACCGTGGATGCGTATGCTAGTATTGACGGTTATATG 103  
DB 375 SerGlyLysLysAspAlaAsn-----GlyArgAlaValLeuValLysGlnGluVal 392  
QY 104 GTCAGCAGCCCTATTTTAAACGTCCTCCCAATCATGCGACGTCTCAGGGAATTT 163  
DB 393 IleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleSer 412  
QY 164 ACCCAGCGTGAAGTGAACAACCTCGCTCAGATTTAAATCTGGAGGATGCTTTTGT 223  
DB 413 AsnProAsnGluAlaArgGlnLeuSerLeuLeuAlaGlyAlaLeuAlaPro 432  
QY 224 CCCGAGGTTCTCAGTGAAGACGATCTCTCTGATCTTGGGAAAAACAATGTACAA 283  
DB 433 IleGlnIleValGluGluArgThrIleGlyProThrLeuGlyMetGlnAsnIleLysGln 452  
QY 284 GGCAATATCTCAGCATGCTGTGGCTTGGCAATGCTATTGTTTGGATGAGCGTATATT 343  
DB 453 GlyLeuGluAlaCysLeuAlaGlyLeuValSerIleLeuPheMetIlePhePheTyr 472  
QY 344 AGATTTGGAGCGTCATCGCTCGGGAGCTGTCTCTCTGAATCTTTTCTTATCTGGCA 403  
DB 473 LysLysPheGlyLeuIleAlaThrSerAlaLeuValAlaAsnLeuValLeuValGly 492  
QY 404 GCTCTACGATTTTG---GATGCCCACTACCTGTGTCAGGACTCGTGGGATGTTCTT 460  
DB 493 IleMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512  
QY 461 GCTATGGGATCGCGTAGTCAAAATGTTCTTCTATTTCGAAAGAAATCCGAGAGGAAT 520  
DB 513 ThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGluArgIleLysGluLeu 532  
QY 521 TTATTGTCTCAAGTCTTAAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCC 580

Db 533 SerAsnGlyArgThrValGlnGlnAlaIleAsnGluGlyTyrAlaGlyAlaPheSer 552  
QY 581 ATTTTGTGATTTAACTTGACTACAGTATTGGCTCAGCACCTCTTTTCTCTAGATACA 640  
Db 553 IlePheAspAlaAsnIleThrThrLeuIleValIleLeuIleValGlyThr 572  
QY 641 GGGCTATTAAAGGGTTTCTTTCGATGATGATTTAGGAATTTCTCTCAATGTTTACG 700  
Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaThrSerMetPheThr 592  
QY 701 GCTCTTTTTCATGACTAA 718  
Db 593 AlaIleIleGlyThrArg 598  
RESULT 15  
ID SECF\_HELPJ STANDARD; PRT; 323 AA.  
AC Q9ZJ65;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Protein-export membrane protein secF.  
GN SECF OR JHP1450.  
OS Helicobacter pylori J99 (Campylobacter pylori J99).  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
OX NCBI\_TaxID=85963;  
RN [1]\_TaxID=85963;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99120557; PubMed=9923682;  
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,  
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,  
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,  
RA Trust T.J.;  
RT "Genomic sequence comparison of two unrelated isolates of the human  
RT gastric pathogen Helicobacter pylori";  
RL Nature 397:176-180(1999).  
CC -!- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).  
CC -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS  
CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY  
CC (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE001567; AAD07023.1; -;  
DR InterPro; IPR003335; SecD\_SecF.  
DR Pfam; PF02355; SecD\_SecF; 1.  
DR TIGRFAMS; TIGR00916; 2A0604s01; 1.  
DR TIGRFAMS; TIGR00966; 3a0501s07; 1.  
KW Protein transport; Translocation; Transmembrane; Membrane;  
KW Complete proteome.  
FT TRANSMEM 19 39 POTENTIAL.  
FT TRANSMEM 138 158 POTENTIAL.  
FT TRANSMEM 162 182 POTENTIAL.  
FT TRANSMEM 244 264 POTENTIAL.  
FT TRANSMEM 269 289 POTENTIAL.  
SQ SEQUENCE 323 AA; 36142 MW; 15363AA8FF0E9215 CRC64;  
  
Alignment Scores:  
Pred. No.: 1,24e-20 Length: 323  
Score: 351.00 Matches: 91  
Percent Similarity: 50.81% Conservative: 66  
Best Local Similarity: 29.45% Mismatches: 104



QY 950 ATCAGCATGTTGCTCAATCGTGGC-----AAAGTTGTGCATAAA 991  
 Db 49 -----SerLeuValGlnValArgPyrThrGlnAenAlaProIleLysLeuValArgAsp 66  
 QY 992 CTACAGGAAGCTGGTCTTCTCTAGAGACTTCGATTCACAACTTTGGATCTTCAGAA 1051  
 Db 67 LeuPheGluLysGluAlaArgPheLysGlyValGlnValSerGluPheGlySerLysGlu 86  
 QY 1052 AAGATCAAAATCTATTATTAGTATAAAGCTTTAAGCTTACTATCAGCAGCAGCCTCT 1111  
 Db 87 GluIleLeuLysPheProPheValGluThrAlaGluAenGluAenGluAenAla--- 105  
 QY 1112 CTCCTAAATTAACGATCATGAGCTGGGCTTATTGTGGATTTGTTCAGAAACAGGCT 1171  
 Db 106 -----IleValAlaAen----- 109  
 QY 1172 AGATTCTCTACGGAACCTCTAAACGAAACGCAAAATTTGGTCAAGAGTAAGCAGCAAA 1231  
 Db 110 ---IleLeuLysProSerGlyAspPheGluLysPheAspThrValGlyProArg 128  
 QY 1232 CTATCGAAGAAATCGGTTATCAGCGCACCATCGGCTTTTAGGAGCTTTTGGCAATCATC 1291  
 Db 129 ValGlySerGluLysGluLysGlyLysLeuSerLeuLysLeuAlaLeuAlaIle 148  
 QY 1292 TTGCTCTATGAGCTTTCGCTTGAATGGCAATATGCTTTCAGTCGCGTATGCGCTTTA 1351  
 Db 149 MetValTyValSerPheArgTyGluTrpArgPheAlaLeuAlaSerValIleAlaLeu 168  
 QY 1352 ATTATGAC-----CTTTTCGCTACCTGTCAGCTCTTGTATTATAGCACATTTCTTTTG 1405  
 Db 169 ValHisAspValIleLeuValAlaSerSerValIleValPhe----- 182  
 QY 1406 AAGAAATTCAAATAGATTTCAGGCCATTCGTTTAAATGACTGTATTGGGTAATTC 1465  
 Db 183 ---LysIleAspMetAsnLeuGluValIleAlaLeuLeuLeuLeuLeuGlyTySer 201  
 QY 1466 TTAACAATATCTGATCATTTTTCATCGTATTCGTGAAGATCGCAAGCAACTGTTT 1525  
 Db 202 IleAsnAspThrIleIlePheAspArgIleArgGluGluMetLeuSerGlnLysThr 221  
 QY 1526 ACCCTATGATGTTTATGTTAATGATCCCTCAAAAGAGCTTCAGCGCACGTAATG 1585  
 Db 222 LysAsnAlaThrGlnAlaIleAspGluAlaIleSerSerThrLeuThrArgThrLeuLeu 241  
 QY 1586 ACAACAGCTACAACCTATCATGTTTGTATGCTTTTGTATAGCGGCTCCTCTGTC 1645  
 Db 242 ThrSerLeuThrValPhePheValValLeuLeuLeuCysValPheGlySerLysIleIle 261  
 QY 1646 TTTAATTGCTATTATATGACCATGAGGATTTCTTAGGAATTTATCTCTCTTTAT 1705  
 Db 262 IleGlyPheSerLeuProMetLeuIleGlyThrIleValGlyThrTySerSerIlePhe 281  
 QY 1706 ATTGACCACTCTGTTGTTGTTATG 1732  
 Db 282 IleAlaProLysValAlaLeuLeuLeu 290

RESULT 17  
 SECF\_RICPR  
 ID SECF\_RICPR STANDARD; PRT: 311 AA.  
 AC Q2E34;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein-export membrane protein secf.  
 GN SECF OR RP14.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsia;  
 OX NCBI\_TaxID=782;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Madrid E;  
 RX MEDLINE=99039499; PubMed=9823893;

RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
 RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria".  
 RL Nature 396:133-140(1998).  
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT.  
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS  
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: AJ235270; CAA14583.1; -;  
 DR InterPro: IPR003335; SecD\_SecF.  
 DR Pfam: PF02355; SecD\_SecF\_1.  
 DR TIGRFAMs: TIGR00916; 2A0604801; 1.  
 DR TIGRFAMs: TIGR00966; 3a0501807; 1.  
 KW Protein transport; Translocation; Transmembrane; Membrane;  
 KW Complete proteome.  
 FT TRANSMEM 36 56 POTENTIAL.  
 FT TRANSMEM 140 160 POTENTIAL.  
 FT TRANSMEM 164 184 POTENTIAL.  
 FT TRANSMEM 194 214 POTENTIAL.  
 FT TRANSMEM 246 266 POTENTIAL.  
 FT TRANSMEM 272 292 POTENTIAL.  
 SQ SEQUENCE 311 AA; 35175 MW; 31863A150F32E630 CRC64;

Alignment Scores:  
 Pred. No.: 6,58e-20 Length: 311  
 Score: 342.00 Matches: 94  
 Percent Similarity: 49.85% Conservative: 72  
 Best Local Similarity: 28.23% Mismatches: 113  
 Query Match: 10.33% Indels: 54  
 DB: 1 Gaps: 13

US-09-868-987-1 (1-1864) x SECF\_RICPR (1-311)

QY 788 ATAAAGCATGATCTCTGAGAGGATGCAAAAACCTTTGGCTGTTCTGGAAGTGTCTT 847  
 Db 13 IleAspPheAspPheMetAsnPheLysLysValSerTySerPheSer-----Ile 29  
 QY 848 CTTTGTAGTTGCGTTGCTCTCGGTTTGGAGCTGGAATTCGGTT-----TTG 895  
 Db 30 IleLeuSerLeuLeuSerLeu-----IleTrpIleSerIleTyLysPheAsnPhe 46  
 QY 896 GGAATGGAATTTAAAGAGGATGATCGCTTTACCTTTAAT----- 934  
 Db 47 GlyIleAspPheValGlyIleValIleGluValArgLeuAspGlnAlaProAspLeu 66  
 QY 935 CCAAAAGAGCATGCGATCAGCGATGTTGCTCAATGCGTGGCAAAAGTTGTGCATAACTA 994  
 Db 67 ProllysMetArgAlaValLeuSerAlaLeuGluIle---GlyGluValVal-----Leu 83  
 QY 995 CAGGAAGCTGCTTTCTTCTAGAGACTTCGGTATTCAACATTTGGATCTTCAGAAAG 1054  
 Db 84 GlnAsnPheGly---SerGluArgAspLeuSerIleArg---PheGlySerSerSerGlu 101  
 QY 1055 ATCAAAATCTATTATTAGTATAAAGCTTTAAGCTTATAGCTATATAAGCAGACGCTCTCTC 1114  
 Db 102 -----GluAsnLeuMetLysAsnIleAspIleIleLysThrSerLeu 115  
 QY 1115 CTAAATAATACGATCATGAGCTGGCGTTATTGTGGGATTGTTGTACAGAAACAGCGCTAGA 1174  
 Db 116 -----ArgAsnAsn 118

QY 1175 TTCTCTACGGAACCTTAAGCAAGCAAAATTTGTCAAGGTACGCAAACTA 1234  
 DB 119 PheProlyrAsnPhelGlyrArgLysValAspPhe-----ValGlyProGlnVal 135  
 QY 1235 TCGAAGAAATGCGCTTATCATGCGGCGCATTCGGCTTTAGAGCTTTTGCAATCTTG 1294  
 DB 136 GlyArgGlnLeuIleGlnAlaGlyAlaMetCalaMetLeuPheSerPheLeuAlaIleMet 155  
 QY 1295 CTCATGTAGTGGCTGGCTTTGGAATGCAATATGCTTCACTGCTTGGCTTAATT 1354  
 DB 156 ValTyrIleGlyValArgPheGlnTyrPheGlyIleLeuIleAlaLeuVal 175  
 QY 1355 CATGACCTTTGGCTACCTGCGAGCTGTTTATGACACATTTCTTTGCAAAAT 1414  
 DB 176 HisAspVal-----IleLeuAlaLeuGlyPheMetSerMetCysLeu 190  
 QY 1415 CAATATGATTTGCAAGCATTTGCTTAACTGATTTGGGCTATTTAAACAT 1474  
 DB 191 AspheAsnLeuSerThrIleAlaAlaValLeuThrIleIleGlyTyrSerValAsnAsp 210  
 QY 1475 ACTTGATCATTTTGTATGCTGATTCGTGAAGATCGCAAGCAACCTTTAACCCATG 1534  
 DB 211 SerValValIleTyrAspArgIleArgGlnAsnLeuArgLysTyrHisLysAsnIle 230  
 QY 1535 CATGTTTATGATTAATGATGCCCTTCAAAAGACCTTCACCGCATATGCAACACCT 1594  
 DB 231 ThrGlnIleIleAsnLeuSerIleAsnGlnTyrLeuSerArgThrIleLeuThrValIle 250  
 QY 1595 ACAACTCATCATGTTTGTATGCTTGTATGCTTGTATGCGGCTCTGCTTTAATTT 1654  
 DB 251 ThrThrLeuLeuAlaAsnLeuAlaLeuLeuLeuPheGlyGlyAlaIleHisSerPhe 270  
 QY 1655 GCATTATATGACCAATAGAGATTTCTTACGAACTTATGCTCTTTAATATGACCA 1714  
 DB 271 SerValLeuValAlaPhePheGlyIleIleAlaGlyTyrSerSerIlePheIleSerAla 290  
 QY 1715 CCTCGTGTG---TTGTTTATGCTCGCTTAAGAAATCGC 1750  
 DB 291 ProIleLeuThrMetPheAlaAsnArgLysPheAsnLys 303  
 RESULT 18  
 SECF\_HAEMIN STANDARD; PRT: 325 AA.  
 AC P44590;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-export membrane protein secf.  
 GN SECF OR HI0239.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 CX NCBI\_Taxid=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 KereJavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 Uetrich J.T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,  
 Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 Venter J.C.,  
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.";  
 RL Science 269:496-512(1995).  
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).  
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS  
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH

CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (By similarity)  
 CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U32710; AAC21907.1; -.  
 DR TIGR: HI0239; -.  
 DR InterPro: IPR003335; SecD\_SecF.  
 DR Pfam: PF02355; SecD\_SecF.1.  
 DR TIGRPFAMs: TIGR00916; 2A0604801; 1.  
 DR TIGRPFAMs: TIGR00966; 3A0501807; 1.  
 KW Protein transport; Translocation; Transmembrane; Inner membrane;  
 KW Complete proteome.  
 FT TRANSMEM 36 56 POTENTIAL.  
 FT TRANSMEM 148 168 POTENTIAL.  
 FT TRANSMEM 187 207 POTENTIAL.  
 FT TRANSMEM 254 274 POTENTIAL.  
 FT TRANSMEM 281 301 POTENTIAL.  
 SO SEQUENCE 325 AA; 36051 MW; EBC59299ED1CF1F2 CRC64;  
 Alignment Scores:  
 Pred. No.: 6,62e-20 Length: 325  
 Score: 342.00 Matches: 98  
 Percent Similarity: 51.90% Conservat: 80  
 Best Local Similarity: 28.57% Mismatches: 111  
 Query Match: 1 Indels: 54  
 DB: 1 Gaps: 15  
 US-09-868-987-1 (1-1864) x SECF\_HAEMIN (1-325)  
 QY 710 ATGACTAATTTTCTTCATGCTGTGATGATTAAGCCCACTACACAGTTGCATATG 769  
 DB 1 MetMetLysLeuPhe-----ThrLysAspLysAspGlyHisPhe 13  
 QY 770 ATGATATAGTTCTGCGGATTAAG-----CATGATTTCTTGAGAGATGC 814  
 DB 14 IleArgIleLeuAsnGlyIleLysLeuProPheProLeuThrGlnPheMetLysValArg 33  
 QY 815 AAAAATCTTGGCGCTGTTTGTGAAGTGTTCCTTTAGCT-----TGGCTTGT 865  
 DB 34 LysLeuGlyTyrIleLeuSerAlaLeuMetValIleSerLeuPhePheIleIleThr 53  
 QY 866 CTGCGTTTGGAGCTTGGAATTCGTTTGGGAATGATTTTAAAGAGGATATGCTT 925  
 DB 54 LysGlyPheAsn--TTP-----GlyLeuAspPheThrGlyValAlaPhe 68  
 QY 926 ACCTTATATCAAAAGAGATGCGATGCGATGCTCAAAATGCGTGCACAAAGTTGTG 985  
 DB 69 AsphThrHisPheSerGlnSer-----AlaAsnLeuGlnIleArgSer----- 83  
 QY 986 CATTAAGTACAGAGAGCTGCTTCTTCTTACGACTTCGCTATTAACATTTGATCT 1045  
 DB 84 ---LysLeuHisGlnAlaGlnGlyIleGlnSerPro-----IleValGlnThrThrGlySer 100  
 QY 1046 TCAGAAAAGATCAAAATCTATTTTATGATTAAGCTTAAAGCTTAACTACTAAGCAGATCGA 1105  
 DB 101 ValGlnAspValMetIle-----ArgLeuPro 109  
 QY 1106 GCGTCTCTCTAAATTAATGATCATGATGCTGGGCTTATGTTGATGTCAGAAAC 1165  
 DB 110 AlaSerAsnAsnAspSerThrIle-----GlyGlnHisValLysSerMetLeuGlnAsn 127  
 QY 1166 AGGCTAGATTTCTCTACGGAATCTTAAAGAAACGCAAAATTTTGGTCAAAAGTAAAC 1225  
 DB 128 ValAspLysAspIle-----GlnIleArgSerIleGlnPhe-----ValGly 141

QY 1226 AGCAAACTATCGAAGAAATGCGTTATCAGGGACCATCGGGCTTTTAGGAGCTTTGGCA 1285  
 DB 142 ProAsnValGlyGluGluAlaGlnGlyAlaValTyAlaThrLeuAlaThrLeuAla 161  
 QY 1286 ATCATCTGCTCTATGCTAGTTGCGCTTCAATGCGCAATATGCTTCAGTGGCGTATGC 1345  
 DB 162 MetValLeuIleTyValGlySerArgPheGluTrpArgLeuGlyPheGlySerIleAla 181  
 QY 1346 GCTTAATTCATGACCTTTTGGCTACCTGTCGAGCTTGTGTTATAGCACATTTCTTTTG 1405  
 DB 182 SerLeuAlaHisAspValIleIleThrLeuGlyVal-----PheSerAla 196  
 QY 1406 AAGAAATTCAAATAGATTGCAAGCCATGCTGTTTAAATGACTGTATGGGATTTCA 1465  
 DB 197 LeuGlnIleGluIleAspLeuThrPheValAlaIleLeuSerValValGlyTySer 216  
 QY 1466 TTAACAATACCTTGATCATTTTGGCTGCTATTCGTGAAGAT---CGCAAGCGAACCTG 1522  
 DB 217 IleAsnAspSerIleValPheAspArgValArgGluAsnPheArgIleAargArg 236  
 QY 1523 TTTACCCCTATGCTATGTTTATGTTAATGATGCCCTTCAAAAGACCTTCAGCGCAGGTA 1582  
 DB 237 LeuAspThrIleAspIle---IleAspIleSerLeuThrGlnThrLeuSerArgThrIle 255  
 QY 1583 ATGCAACAGCTACACTCTATCATGTTTGTATGCTTTTGTATAGCGGCTCTCTCT 1642  
 DB 256 IleThrSerValThrThrLeuValValMetAlaLeuPhePheGlyGlyProSer 275  
 QY 1643 GTCCTTAATTTGCTATTTATGACCATAGGATTCCTTAGGAACTTTATGCTCTCT 1702  
 DB 276 IleHisAsnPheSerLeuAlaLeuValGlyIleGlyPheGlyThrTySerSerIle 295  
 QY 1703 TATATTGCA 1711  
 DB 296 PheValAla 298  
 RESULT 19  
 SECD BORBU  
 ID SECD BORBU STANDARD; PRT; 586 AA.  
 AC OS1596;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein-export membrane protein secD.  
 GN SECD OR BB0652.  
 OS Borrelia burgdorferi (Lyme disease spirochete).  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxId=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35210 / B31;  
 RC MEDLINE=98065943; PubMed=9403685;  
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,  
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,  
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,  
 RA Peterson J., Kervilange A.R., Quackenbush J., Salzberg S., Hanson M.,  
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
 RA Uterback T., Wathey L., McDonald L., Artiach P., Bowman C.,  
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
 RA Smith H.O., Venter J.C.;  
 RT "Genomic sequence of a Lyme disease spirochaete, Borrelia  
 burgdorferi".  
 RL Nature 390:580-586(1997).  
 CC -!- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).  
 CC -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS  
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH  
 CC (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.ebi.ch/announce/  
 CC or send an email to license@ebi-sib.ch).  
 CC -----  
 CC EMBL; AB001166; AAC66993.1; -.  
 DR TIGR; BB0652; -.  
 DR InterPro; IPR003335; SecD SecF.  
 DR Pfam; PF02335; SecD SecF; 1.  
 DR TIGRFAMs; TIGR00916; 2A0604s01; 1.  
 DR TIGRFAMs; TIGR01129; secD; 1.  
 KW Protein transport; Translocation; Transmembrane; Membrane;  
 KW Complete proteome.  
 FT TRANSMEM 7 27 POTENTIAL.  
 FT TRANSMEM 418 438 POTENTIAL.  
 FT TRANSMEM 439 459 POTENTIAL.  
 FT TRANSMEM 521 541 POTENTIAL.  
 FT TRANSMEM 546 566 POTENTIAL.  
 SQ SEQUENCE 586 AA; 65023 MW; DDCD08EC4B47FBC CRC64;  
 Alignment Scores:  
 Pred. No.: 3 43e-19 Length: 586  
 Score: 333.50 Matches: 80  
 Percent Similarity: 55.92% Conservative: 57  
 Best Local Similarity: 32.65% Mismatches: 103  
 Query Match: 10.07% Indels: 5  
 DB: 1 Gaps: 4  
 US-09-868-987-1 (1-1864) x SECD BORBU (1-586)  
 QY 62 AACGCTGGATGGCGATGGCTAGTGTAGTGTATGATGACGGTATATGTCACAGCCCTATTTTA 121  
 DB 338 AsnValGlySerLeuAlaValMetGluGlyLysIleLysSerValAlaGlyIle 357  
 QY 122 AAGTCCCATTTGAAA---AATCATGCCAGTGTCTCAGG---AAATTTACCCACCGTGAA 175  
 DB 358 GlyTyAlaIleThrGlyAsnValSerIleGlnGlyAspSerPheAspLysLysGlu 377  
 QY 176 GTGAGCAAACTCGCTCAGATTTAAATCTGGAGCATGCTTTTGTCCGAGGTTCTC 235  
 DB 378 AlaGlnAspLeuAlaLeuValPheLysThrAlaAlaPheProValAspIleLysIleAsp 397  
 QY 236 AGTGAAGAGACGATCTCTCTGATCTTTGGGAAAAACAATGTACACAGGCAATTTCTCA 295  
 DB 398 AspLeuArgIleIleGlyProThrLeuGlyAlaArgThrIleAspLeuGlyIleLysAla 417  
 QY 296 GCATGCTGCTGGCAATGCTTATTTGTTTGTAGTACGCTATATATAGATTTCGAGGC 355  
 DB 418 SerAlaLeuAlaLeuCysLeuValPheIleCysValTyTyGlyLeuSerGly 437  
 QY 356 GTCATCGCTTCG---GGAGCTGTTCTTCTGAATCTTTTGTCTATCTGGCAGCTTACAG 412  
 DB 438 ValValAlaGlyPheSerLeuValIleTyAsnValPheLeuIleLeuAlaIleLeuSer 457  
 QY 413 TATTGTGATCGCCACCTCACCTGCTCAGGACTCGCTGGGATTTGTTTGTATGGGATG 472  
 DB 458 AlaPheAsnPheThrLeuThrLeuThrSerIleAlaGlyLeuIleLeuThrMetGlyMet 477  
 QY 473 GCGCTAGATGCAATGTTCTTGTATTCGAAAGAAATCCGAGAGGAATTTTATTGCTCTCAA 532  
 DB 478 AlaValAspIleAsnIleValIleTyGluArgIleLysGluGluIleArgGluGlyArg 497  
 QY 533 AGTCTTAAAAAATCTGTGAAGAAAGGATATACCAAGCTTTTGGAGCCATTTTGTATCT 592  
 DB 498 ArgPheGluAsnAlaPheGluAlaGlyPheLysAlaPheLeuSerIleMetAspAla 517  
 QY 593 AACTGTACTACAGTATGGCTCAGCACTCTCTTCTCTCTAGATACAGGCGCTATTAAA 652  
 DB 518 AsnIleThrThrPheIleAlaValLeuPheLeuThrLeuLeuGlyThrGlyValIleGln 537  
 QY 653 GGGTTCCTTTGACATGATTTTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATG 712

```

Db 538 GlyheAaTatpserLeuSerValGlyIleValAlaSerLeuPheSerLeuIlePhe 557
Qy 713 ACTAAATTTTC-----TTCATGCTGTGATGAATTAAGACCAACATACAGTTGCAT 766
Db 558 SerArgPheIleLeuGluPheIleIleSerValArgLysSerLysPheIleSerIleSer 577
Qy 767 ATGATGATTAAGTTC 781
Db 578 TrpSerSerLysTyr 582

RESULT 20
SECF_AQUAE STANDARD; PRT; 288 AA.
ID SECF_AQUAE STANDARD; PRT; 288 AA.
AC 067536;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein-export membrane protein secf.
GN SECF OR AO 1602.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RA MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aubay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 352:353-358(1998).
CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS
CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (by similarity).
CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC EMBL: AE000747; AAC07496.1; -
CC InterPro: IPR003335; SecD_SecF.
CC DR Pfam: PF02355; SecD_SecF; 1.
CC DR TIGRFAMs: TIGR00916; 2A0604s01; 1.
CC DR TIGRFAMs: TIGR00966; 3A0501s07; 1.
CC KM Protein transport; Translocation; Transmembrane; Complete proteome.
CC
CC FT TRANSMEM 13 POTENTIAL.
CC FT TRANSMEM 143 POTENTIAL.
CC FT TRANSMEM 145 POTENTIAL.
CC FT TRANSMEM 175 POTENTIAL.
CC FT TRANSMEM 240 POTENTIAL.
CC SQ SEQUENCE 288 AA; 31632 MW; E526B8D5BBD99770 CRC64;

Alignment Scores:
Pred. No.: 1.42e-17 Length: 288
Score: 313.00 Matches: 93
Percent Similarity: 51.08% Conservative: 72
Best Local Similarity: 28.79% Mismatches: 110
Query Match: 9.45% Indels: 48
DB: 1 Gaps: 12

US-09-868-987-1 (1-1864) x SECF_AQUAE (1-288)
Qy 800 TTCTTGAGAGATGCAAAACTTTGGGCTTTCTGGAAGTGTCTTCTTTAGGTTGC 859
|||||
|||
:::
|||||
:::
|||||
:::

```

```

Db 6 PheLeuArgLeuArgLysAlaIleTyrGlyValSerAlaLeuLeuIleSerLeu 25
Qy 860 GTTGCTCTCGGGTTTGGAGCCTTGAAATCCGTTTGGGAATGATTTTAAAGAGG--- 916
Db 26 LeuSerLeuLeuPheArgGlyLeuAsn-----LeuGlyLeuAspPheThrGlyGlyThr 43
Qy 917 ---TATGCTTTACCTTTATATCAAAAGACATGACATGACGATGCTGCTCAAAATGCT 973
Db 44 LeuTyrGluValLysPheGlyLysSer-----ValAspIle----- 55
Qy 974 GGCAGAGTTGTGATTAACATGACAGAGAGCTGCTTTCTTCTGAGACTTCCTGATTCAA 1033
Db 56 GlyLysLeuArgLysThrIleSerSerIleGlyIle-----LysGlyPheLeuIleGln 73
Qy 1034 ---ACATTGATCTTCAGAAAAGATCAAAATCTATTGATGATTAAGCTTAAGCTAT 1090
Db 74 GluThrLysGluGlyThrPheValIleLysVal-----LysThrGlyGluPro 89
Qy 1091 ACTAAGCAGATAGAGGCTCTCTCTAAATTAACGATCATGAGCTGCGCTTATGTGGG 1150
Db 90 ValGluLysLeuGluLysValLysLys-----ValAspIle----- 99
Qy 1151 ATTGTTGTCAAAAACAGGCTTATTTCTCTACGAAACTCTAAGCAAAACGCAAAATTT 1210
Db 100 -----PheGlyLysTyrGluLeuLeuIleArgLys--- 108
Qy 1211 TGGTCAAGATTAAGCAAACTATGAGAAAATGCGTTATAGGCGACATCGGCGTT 1270
Db 109 ---GluThrLysSerGlyValLysSerGluGluLeuGlnLysAlaValAlaPheAlaIle 127
Qy 1271 TTGAGAGCTTGGCAATCACTGCTGCTATGAGAGTTGAGAGTTGAGAGTTGAGAGTTGCT 1330
Db 128 LeuThrAlaLeuGlyLysIleLeuLeuTyrLysGlyValArgPheGlnProValIlePheGly 147
Qy 1331 TTGAGTCCGTATGCGCTTATTAATGACCTTTGGCTACCTGTGAGCTGTTGTTATA 1390
Db 148 PheGlyValAlaIleLeuAlaLeuAlaHisAspValIleThrVal-----Leu 162
Qy 1391 GCACATTTCTTTTGAAGAAAATCAATGATTTGCAAGCCATTTGGTCTTTAATGACT 1450
Db 163 GlyAlaTyrSerIleThrGlnArgGluValAsnLeuGluValAlaSerAlaIleLeuVal 182
Qy 1451 GATTTGGGATTCATTAACAACTATGATGATTTGATTTGATTTGATTTGATTTGATTTGAT 1507
Db 183 ValAlaGlyTyrSerValAlaAspThrValIlePheAspArgIleArgGluAsnLeu 202
Qy 1508 CGCCAGACGAACCTGTTTACCCCTATGACATGTTTATGATGATGATGATGATGATGATG 1567
Db 203 ArgLysLysLysGlyPheThr-----LeuGluGluIleMetAspLeuSerIleAsnGlnThr 221
Qy 1568 TTACGCCGACGCGTATGACACAGCTACATCTTATGATTTGTTATGCTTTGTTGTT 1627
Db 222 LeuSerArgThrIleMetThrSerLeuThrThrLeuValThrAlaLeuThrLeuPheIle 241
Qy 1628 ATAGGGGGCGCTGCTGCTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 1687
Db 242 LeuGlyGlyLysAlaLeuSerAsnIleMetPheAlaPheValAlaGlyValValGly 261
Qy 1688 ACTTATGCTCTCTTATATGACACACCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1747
Db 262 ThrTyrSerSerValPheValAlaSerAlaPheValLeuAspMetGlnLysLeuPheLys 281
Qy 1748 CGCTCAAAA 1756
Db 282 ArgGlyGlu 284

RESULT 21
SECD_MYCLE STANDARD; PRT; 571 AA.
ID SECD_MYCLE STANDARD; PRT; 571 AA.
AC P38387; O69485; O9CCT3;
DT 01-OCT-1994 (Rel. 30, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```





RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Salston J.E., Taylor K., Whitehead S., Barrett B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT.  
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLATION APPARATUS  
 CC WHICH COMPRISE SECF, SECB, SECD, SECE, SECF, SECG AND SECY  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL: Z77724, CAB01256.1;  
 DR EMBL: AE007100, AAK46977.1; ALT\_INIT.  
 DR TIGR: MT2664;  
 DR TubercuList; RV2587C;  
 DR InterPro: IPR00335; SecD\_SecF.  
 DR Pfam: PF02355; SecD\_SecF; 1.  
 DR TIGRFAMs: TIGR00916; 2A0604s01; 1.  
 DR TIGRFAMs: TIGR01129; secD; 1.  
 KW Protein transport; Translocation; Transmembrane; Membrane;  
 KM Complete proteome.  
 FT TRANSMEM 13  
 FT TRANSMEM 385 405 POTENTIAL.  
 FT TRANSMEM 410 430 POTENTIAL.  
 FT TRANSMEM 441 461 POTENTIAL.  
 FT TRANSMEM 489 509 POTENTIAL.  
 FT TRANSMEM 514 534 POTENTIAL.  
 SQ SEQUENCE 573 AA; 60:67 MW; 5A8F42E2C0389CE2 CRC64;

Alignment Scores:  
 Pred. No.: 9.83e-17 Length: 573  
 Score: 303.00 Matches: 78  
 Percent Similarity: 51.13% Conservative: 58  
 Best Local Similarity: 29.32% Mismatches: 100  
 Query Match: 9.15% Indels: 30  
 DB: 1 Gaps: 6

US-09-868-987-1 (1-1864) x SECD\_MYCTU (1-573)

QY 5 ACTTCGCAATTTGTCAGAGGATGACGGGCACTGCTAATGCAATAT----- 55  
 DB 277 Thirsergiymetaspiglnargllylellyrvalvalasplelgnphelyselgylpro 296  
 QY 56 ----- -CTGCAACCGCTGATGCGGATGCGGATGCGGTATG 88  
 DB 297 Alaalaaniletrpalasptyrthralaahstlellymrglnthralaherthr 316  
 QY 89 ATTGACGGTTATGTTAGACAGCCCTATTAAACGTCCATGTA---AATCATGCC 145  
 DB 317 Leuapserglnvalvaliserlalerprolnlelgnlualallelprogllygllyargthr 336

QY 146 AGTGTCTCAGG-----AAATTACCCAGCGTGAAGTGAACAACTCGCTCAGAT 196  
 DB 337 Glnleiserllyglasppropophemthralalathralaarglnleualasval 356  
 QY 197 TTAATCTGAGAGCATGTCCTTTGTTCCGAGTTCTCAGTGAAGACATCTCTTC 256  
 DB 357 Leuylserlylserleuproleuserphedlnproserclualaglnthralaserla 376  
 QY 257 GATCTTGGAAAAAACAATATACACAGGCAATATCTCAGCATGCTGTGGCATG 316  
 DB 377 ThirleuillyneuserserleualaglylmecllalacllylallellyleuLeu 396  
 QY 317 CTATGTTTGTGATGACGGATATATTAGATTGGAGCC-----GTC 358  
 DB 397 ValleunallylserleuLeuYrYrYrYrValleunlyleuLeuThralaleuSerleu 416  
 QY ATCGCTCGGAGAGCTGTCCTTGAATCTTTGCTTATCTGCGGACGCTACAGTATTG 418  
 DB 417 ValalaserllyserleuValpnealalleleuValleunly-----ArgTyrile 434  
 QY 419 GATCGCCACTACCTGTCAGACCTCGCGATGTTCTGTATGCGGATGCCGTA 478  
 DB 435 Asnlyrthirleuaspneualaglyllealaglyleuilellellyllythrtrala 454  
 QY 479 GATGCAATGTCCTGATTCGAAAGATCCGAGGATTTTATGTCCTCAAGCTT 538  
 DB 455 Aspserhevalvalpnegluarglyleuasplulleargluglyargserphe 474  
 QY 539 AAAAAATCTGTAGAAAAAGATATACCAAGCTTTTGAAGCCATTTTGAATTAATTG 598  
 DB 475 Argserlavalproarglytrpalalarglalythrlevalsergllyasnal 494  
 QY 599 ACTACGATTTGGCTCAGACCTTTCTTCTTGAATACAGGCGCTATTAAGGTTT 658  
 DB 495 ValthrheleuallaalavalleuYrYrYrYrPheleualallellylgnllylsglyphe 514  
 QY 659 GCTTTCATGATTTTAGAATTTCTCTTCATGTTTACGGCTCTTTCAGTACGATTA 718  
 DB 515 Alapherthleu-----GlyleuThrthirleuaspneualval 528  
 QY 719 TTTTCTTCATGCTGTGG 736  
 DB 529 Valpheuvalthrtrp 534

RESULT 23  
 SECF BORBU STANDARD; PRT; 299 AA.  
 AC 051597;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-export membrane protein secF.  
 GN SECF OR BB0653.  
 OS Borrelia burgdorferi (Lyme disease spirochete).  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35210 / B31;  
 RX MEDLINE=98065943; PubMed=9403685;  
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,  
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,  
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,  
 RA Peterson J., Kariavage A.R., Quackenbush J., Salzberg S., Hanson M.,  
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
 RA Utterback T., Matthey L., McDonald L., Attiach P., Bowman C.,  
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
 RA Smith H.O., Venter J.C.;  
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia  
 RT burgdorferi";  
 RL Nature 390:580-586(1997).  
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).  
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLATION APPARATUS





Db 309 AlaLeuMetVal1AlaValTrpLeuLeuGlyValGlyThrLeuLysAspLeuAlaLeu 328

Qy 1661 ATTAGACCAATGAGGATCTCTCTAGAACCTTAACGCTCTTATATGACCAACCTGTG 1720  
:::|||||  
Db 329 ValGlnLeuIleGlyIleIleIleGlyThrTyrsSerIlePhePheAlaThrProLeu 348  
Qy 1721 TTGTGTGTTT-----ATGGTCGGTAAGAAATCCG 1750  
||||:|||||  
Db 349 LeuValThrLeuArgGluArgThrGlnLeuValArgAsnHisThrArg 364

RESULT 25

SECID	STRCO	STANDARD;	PRT;	570 AA.
AC	053955:	09L293;		
DT	01-NOV-1997 (Rel. 35, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Protein-export membrane protein secD.			
GN	SECID OR SCO1516 OR SCIL2.06C.			
OS	Streptomyces coelicolor.			
OC	Bacteria; Actinobacteria; Actinobacteridae (class); Actinobacteridae;			
OX	NCBI_TaxId=1902;			
RA	11			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A3(2);			
RA	Loriaux A., Frate P., Brans A., Duesert J.,			
RL	Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.			
RP	121			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A3(2) / M145;			
RX	MEDLINE=21996410; PubMed=12000953;			
RA	Thomson S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,			
RA	Bentley S.R., James K.D., Harris D.B., Quail M.A., Kleser H.,			
RA	Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,			
RA	Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,			
RA	Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,			
RA	Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,			
RA	Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,			
RA	Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,			
RA	Hopwood D.A.;			
RT	"Complete genome sequence of the model actinomycete Streptomyces			
RL	coelicolor A3(2)." /			
RL	Nature 417:141-147(2002).			
CC	-1- FUNCTION: INVOLVED IN PROTEIN EXPORT.			
CC	-1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS			
CC	WHICH COMPRESSES SECY, SECD, SECE, SECF, SECG AND SECY			
CC	(BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).			
CC	-1- SIMILARITY: BELONGS TO THE SECY/SECF FAMILY. SECY FAMILY.			
CC	-----			
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; X85969; CA559954.1; .			
DR	EMBL; AL137778; CAB70918.1; .			
DR	InterPro; IPR001036; ACriIvln-res.			
DR	InterPro; IPR003335; SecD SecF.			
DR	Pfam; PF02355; SecD SecF; 1.			
DR	PRINTS; PR00702; ACRIPLAVINRP.			
DR	TIGRFAMs; TIGR00916; 2A0604s01; 1.			
DR	TIGRFAMs; TIGR01129; secD; 1.			
KM	Protein transport; Translocation; Transmembrane; Membrane;			
KM	Complete proteome.			
FT	TRANSMEM 370 390 POTENTIAL.			
FT	TRANSMEM 395 415 POTENTIAL.			
FT	TRANSMEM 419 439 POTENTIAL.			
FT	TRANSMEM 474 494 POTENTIAL.			

```

FT TRANSMEM 498 518 POTENTIAL.
FT CONFLICT 72 72 Y -> N (IN REF. 1).
FT CONFLICT 191 191 S -> K (IN REF. 1).
FT CONFLICT 550 551 PL -> SV (IN REF. 1).
SQ SEQUENCE 570 AA; 59511 MW; 073ADE27F2B921E3 CRC64;

Alignment Scores:
Pred. NO.: 1.34e-14 Length: 570
Score: 276.50 Matches: 68
Percent Similarity: 54.31% Conservative: 58
Best Local Similarity: 29.31% Mismatches: 91
Query Match: 8.33% Indels: 15
DB: 1 Gaps: 4

US-09-868-987-1 (1-1864) x SEC2_STRCO (1-570)
Qy 74 CQTATGGCTGTAGTATGACGGTTATATGTCAGACGCCCTATTATAACGTCCTCATG 133
Db 300 GlnPheAlaIleValLeuAspAsnGluValSerAspProTyrValSerGlnAlaLeu 319
Qy 134 AAA---AATCATGCCAGTGTCTCAGGGAATTTACCCACCGTGAAGTGAGCAAACTCGCC 190
Db 320 ThrGlyGlyAsnAlaGluIleSerGlySerPheAspGlnGluAlaGlnSerLeuAla 339
Qy 191 TCAGATTTAAATCTGGAGCGATGCTTTTGTCCGAGGTTCTCAGTCAGAGAGAGATC 250
Db 340 AsnMetLeuSerTyrGlyAlaLeuProLeuThrPheLysGluAspSerValThrVal 359
Qy 251 TCTCTGATCTGGGAAAAACAATGTACACAAGGCATTATCTCAGCATGCTGTGGCTTG 310
Db 360 ThrAlaAlaLeuGlyGlyGlnLeuLysAlaGlyLeuIleAlaGlyAlaIleGlyLeu 379
Qy 311 GCAATGCTTATTGTTTGTAGAGCGTATATATATAGATTGTGGAGGCGTCATCGCTCGGGA 370
Db 380 AlaLeuValValLeuTyrLeuLeuPheTyrArgGlyLeuSerPheIleAlaValCys 399
Qy 371 GCTGTTCTTGAATCTTTGTTGTTATCTGGGAGCTCTACAGTATTGGATGCG----- 424
Db 400 SerLeuLeuValSerAlaGlyLeuThrTyrValIleMetAlaLeuLeuGlyProThrIle 419
Qy 425 -----CCACTCACCTCTCAGGACTCGCTGGATGTCTTGTCTATGGGATGCCGTA 478
Db 420 GlyPheAlaLeuAsnLeuProAlaValCysGlyAlaIleValAlaIleGlyIleThrAla 439
Qy 479 GATGCAAAATGTTCTGTATTCGAAAGAAATCCGAGAGAAATTTTATTGCTCAAAAGTCTT 538
Db 440 AspSerPheIleValTyrPheGluArgValArgAspGluIleArgGluGlyArgThrLeu 459
Qy 539 AAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTATCTAACTTG 598
Db 460 ArgProAlaValGluArgAlaTrpProArgAlaArgThrIleLeuValSerAspPhe 479
Qy 599 ACTACAGATTGGCCCTCAGCACTTCTTCTCTCCTAGATACAGGCGCTATTAAAGGGTTT 658
Db 480 ValSerPheLeuAlaAlaValLeuPheIleValThrValGlyLysValGlnGlyPhe 499
Qy 659 GCTTTGCATCTG-----ATTTTAGGAATTTTCTCTTCA 691
Db 500 AlaPheThrLeuGlyLeuThrThrLeuLeuAspValValValPheLeuPheThrLys 519
Qy 692 ---ATGTTTACGGCTCTTTTCATGACTAAATTTTTC 724
Db 520 ProLeuLeuThrLeuMetAlaArgArgLysPhePhe 531

RESULT 26
SECF_TREPA
ID SEC2_TREPA STANDARD; PRT; 420 AA.
AC O83426;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-export membrane protein secf.
GN SEC2 OR TP0411.

```

```

OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Winn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT.
CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
CC WHICH COMPRISE SECA, SECB, SECD, SECF, SECG AND SECH
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SEC2/SECF FAMILY. SEC2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; AE001219; AAC65399.1; -.
CC TIGR; TP0411.
CC InterPro; IPR003335; Sec2_SecF.
CC Pfam; PF02335; Sec2_SecF; 1.
CC TIGRFAMs; TIGR00916; 2A0604801; 1.
CC TIGRFAMs; TIGR00966; 3A0501807; 1.
CC Protein transport; Translocation; Transmembrane; Membrane;
CC Complete proteome.
CC TRANSMEM 7 27 POTENTIAL.
CC TRANSMEM 250 270 POTENTIAL.
CC TRANSMEM 276 296 POTENTIAL.
CC TRANSMEM 297 317 POTENTIAL.
CC TRANSMEM 358 378 POTENTIAL.
CC TRANSMEM 388 408 POTENTIAL.
CC SEQUENCE 420 AA; 45731 MW; D7FC706AF596F6F1 CRC64;

Alignment Scores:
Pred. NO.: 4.76e-14 Length: 420
Score: 269.50 Matches: 67
Percent Similarity: 60.59% Conservative: 36
Best Local Similarity: 39.41% Mismatches: 50
Query Match: 8.14% Indels: 17
DB: 1 Gaps: 6

US-09-868-987-1 (1-1864) x SEC2_TREPA (1-420)
Qy 1220 GTAGCGCAAACTATCGAAGAAATCGGTTATCAGCGCACCATCGGCTTTTAGGAGCT 1279
Db 243 ValGlySerArgPheSerHisLeuValArgGlnAlaLeuLeuValLeuGlyAla 262
Qy 1280 TTGGCAATCATCTTGTCTATGTCAGTTTGGCTTGAATGCGCAATATGCTTTCAGTGCC 1339
Db 263 LeuValLeuIlePheLeuTyrValAlaLeuArgPheArgPheAlaLeuGlyAla 282
Qy 1340 GTATGCGCTTTAATTCATGACCTTTTGGCTACCTGTGTCAGCTCTTTTATAGCACATTC 1399
Db 283 IleValAlaLeuValHisAsp-----AlaCysIleMetValSerPheMetValTrpPhe 300
Qy 1400 TTTTGTGAAGAAATTCAAATAGATTTCAGAGCCATTGGTGTCTTAATGACTGTATTGGGG 1459
Db 301 GlyLeu-----GluPheAsnSerAlaSerIleAlaAlaIleLeuThrIleIleGly 317

```

QY 1460 TATTCATTAAACAACTTGTGATGTTGATGCTATGCGTGA-----GAT 1507  
 |||||  
 Db 318 TysTerileuasnaphThValValPheaspArgValArgInThrLeuLeuasp 337  
 |||||  
 QY 1508 CGCCAGCGAACCCTGTTTACCCCTATGATGTTTATGTTATGATGCCCTTCAAAAGACG 1567  
 |||||  
 Db 338 ProtlelaeserValInThrValLeuaspArgSerGlnThrapMleu----- 354  
 |||||  
 QY 1568 TTCAGCGGACGCTATGACAAAGCTATGCTATGCTATGCTTGTATGCTTTGTTT 1627  
 |||||  
 Db 355 ---TharTgThrValValThrThrValThrThrLeuValAlaLeuMleuTyrVal 373  
 |||||  
 QY 1628 ATA-----GGCGGCTCTCTGCTTTTATTTGCTATTTATGACATGAGGATCTT 1681  
 |||||  
 Db 374 PheThrgluTgylGlySerArgaspPheSerLeuAla-----LeuMetValGlyMetVal 391  
 |||||  
 QY 1682 CTAGAACTTATGCTGCTCTTATTTATGCA 1711  
 |||||  
 Db 392 SerGlyValTyrSerThrIleTyrIleAla 401  
 |||||  
 RESULT 27  
 SECF\_SYNY3 STANDARD; PRT; 315 AA.  
 ID SECF\_SYNY3  
 AC 055611;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Protein-export membrane protein secf.  
 GN SECF OR SLR0775.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 NCBI\_taxid=1148;  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96127529; PubMed=8590279;  
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
 RA Sugita M., Tabata S.,  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
 RT region from map positions 644 to 924 of the genome.",  
 RL DNA Res. 2:153-166(1995).  
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT.  
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS  
 CC WHICH COMPRISE SECF, SECB, SECD, SECE, SECF, SECG AND SECY  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL: D64000; BAA10119.1; -  
 DR InterPro: IPR003335; Secd SeCF.  
 DR Pfam: PF02355; Secd SeCF.1.  
 DR TIGR: TIGR00916\_2A0604801; 1.  
 DR TIGR: TIGR00966; 3A0501807; 1.  
 KW Protein transport; Translocation; Transmembrane; Membrane;  
 KM Complete proteome.  
 FT TRANSMEM 12 32 POTENTIAL.  
 FT TRANSMEM 136 156 POTENTIAL.  
 FT TRANSMEM 166 186 POTENTIAL.  
 FT TRANSMEM 188 208 POTENTIAL.  
 FT TRANSMEM 247 267 POTENTIAL.  
 FT TRANSMEM 271 291 POTENTIAL.  
 SQ SEQUENCE 315 AA; 34667 MW; C976449EABE6A35C CRC64;

Alignment Scores:

Pred. No.: 4,68e-13 Length: 315  
 Score: 257.00 Matches: 84  
 Percent Similarity: 43.14% Conservative: 67  
 Best Local Similarity: 24.00% Mismatches: 141  
 Query Match: 7.76% Indels: 58  
 DB: Gaps: 9  
 US-09-868-987-1 (1-1864) x SECF\_SYNY3 (1-315)  
 QY 788 ATAAAGCATGTTTCTTGAAGAGTGCMAAAACCTTGGCTGTTCTGMACTGTTT 847  
 |||||  
 Db 1 MetyleuaspPheuPheuTyrTgIuLysProAlaTyrIleValSerLeuLeuVal 20  
 |||||  
 QY 848 CTTTGAAGTTCGCTGCTCGGCTTGAAGCTCGAATTCGCTTTG----- 895  
 |||||  
 Db 21 LeuIleSerIlePheAlaMetAlaIle---SerTyrAlaGlnPheGlnAlaPropheArg 39  
 |||||  
 QY 896 ---GGAATGATTTTAAAGAGGTATGCTTACCTTTAAT----- 934  
 |||||  
 Db 40 ProGlyLeuaspPheValGlyGlyThrArgLeuGlnLeuGlnLeuGlySalasSer 59  
 |||||  
 QY 935 -----CCAAAGACATGACATCAAGATGTTGCTCAATGCGTGCMAAGTTGTG 985  
 |||||  
 Db 60 AsnAsnCysProAlaAlaIleAspValAlaGluValGlnAspIleuGlyVal---- 78  
 |||||  
 QY 986 CATTAACCTACAGAGACGTGCTCTTCTTCTAGACACTTCGATTTCAACATTGGATCT 1045  
 |||||  
 Db 79 -----GlyLeuGlyAsnSerSerValGlnVal----- 87  
 |||||  
 QY 1046 TCAGAAAGATCAAAATCTATTAGTAAAGCTTAAAGCTATGCTAAGCATATGCA 1105  
 |||||  
 Db 88 -----IleGlnAspTyrThrLeuSerIleArg 96  
 |||||  
 QY 1106 GCCTCTCTCTTAAATTAACGATCATGACGTGCGTATTGTGGATTTGTCAGAAAC 1165  
 |||||  
 Db 97 GlnGlnThrLeuasp-----ValGluGln 104  
 |||||  
 QY 1166 AGCCCTAGATTTCTTACGGAACCTTAAACGCAAAATTT----- 1210  
 |||||  
 Db 105 ArgGluAlaValGlnLysAlaLeuaspGlnGlyIleGlyLysPheaspProGlnThrIle 124  
 |||||  
 QY 1211 ---TGTCAAGAGTAAAGCAAGCAACTATGCAAGAAATAGCGTATACAGCGACATCGG 1267  
 |||||  
 Db 125 GlnIleAspThrValGlyProThrValGlyLysAlaLeuPheArgSerGlyValLeuAla 144  
 |||||  
 Db 145 LeuValIleSerLeuLeuGlyIleIleIleTyrLeuThrIleArgPheGlnLeuAspTyr 164  
 |||||  
 QY 1328 GCTTCAAGCGCGATGCGCTTATTCATGACCTTTGGCTACCTGTGCAAGCTTTGTT 1387  
 |||||  
 Db 165 AlaValPheAlaIleIleAlaLeuLeuTyrAspAlaLeuIleThrMetGlyAla---Phe 183  
 |||||  
 QY 1388 ATGACACATTTCTTTTGAAGAAATTAATGATGATTCAGAGCCATGGTGCTTTAAG 1447  
 |||||  
 Db 184 AlaIlePheGlyLeuValGlyValGluValAspSerLeuPheLeuValAlaLeuLeu 203  
 |||||  
 QY 1448 ACTGATTTGGGTATTCATTAACAACTATGCTATGCTATGCTATGCTGTAAGAT 1507  
 |||||  
 Db 204 ThrIleIleGlyPheSerValAsnAspThrValIleTyrAspArgValArgGlnThr 223  
 |||||  
 QY 1508 CGCCAGCGAACCCTGTTTACCCCTATGACATGTTTATGTTAATGATGCCCTTCAAAAGACG 1567  
 |||||  
 Db 224 LeuGluArgIleSerAspTyrPheAlaAsnAlaValAspAlaValAsnGlnThr 243  
 |||||  
 QY 1568 TTCAGCGGACGCTATGACAAAGCTATGCTATGCTATGCTTGTATGCTTTGTTT 1627  
 |||||  
 Db 244 LeuThrArgSerIleAsnThrSerLeuThrThrSerLeuProLeuValAlaIlePheLeu 263  
 |||||  
 QY 1628 ATAGCGGCTCTGCTGCTTATTTATGCTATTTATGACATGAGGATTTCTTGAAG 1687  
 |||||  
 Db 264 PheGlyGlyAspSerLeuLysPhePheAlaLeuAlaLeuIleIleGlyPheAlaSerGly 283  
 |||||

Qy 1688 ACTTATCGTCTCTTATATATGACACCTCTGTTGTTGTTTATGTCGCGTAAGAAAT 1747  
 Db 284 ValTyrSerSerIlePheMetAlaThrThrLeuTrpAlaTrpTrp---ArgLysTrpArg 302  
 Qy 1748 CGCTCAAAATAGTACCGTTAACTTAATC 1777  
 Db 303 SerProLysAsnProProArgGluMetVal 312

RESULT 28  
 SECF\_STRCO STANDARD; PRT; 373 AA.  
 AC Q53956;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-JUN-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-export membrane protein secf.  
 GN SECF OR SCO1515 OR SCL2.05C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Loriaux A., Frere P., Brans A., Dusart J.;  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Corden-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).

CC -!- FUNCTION: INVOLVED IN PROTEIN EXPORT.  
 CC -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS  
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY  
 CC (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X85969; CAA59955.1; -  
 DR EMBL; ALJ37778; CAB70917.1; -  
 DR InterPro; IPR003335; Secd\_Secf.  
 DR Pfam; PF02355; Secd\_Secf; 1.  
 DR TIGRFAMs; TIGR00916; 2A0604s01; 1.  
 DR TIGRFAMs; TIGR00966; 3A0501s07; 1.  
 KW Protein transport; Translocation; Transmembrane; Membrane;  
 KW Complete proteins.  
 FT TRANSMEM 26 46 POTENTIAL.  
 FT TRANSMEM 142 162 POTENTIAL.  
 FT TRANSMEM 166 186 POTENTIAL.  
 FT TRANSMEM 193 213 POTENTIAL.  
 FT TRANSMEM 251 271 POTENTIAL.  
 FT TRANSMEM 280 300 POTENTIAL.  
 FT TRANSMEM 373 AA; 39710 MW; 28C55B23C62F82ED CRC64;  
 SQ SEQUENCE

Alignment Scores:  
 Pred. No.: 5,75e-13 Length: 373  
 Score: 256.00 Matches: 86  
 Percent Similarity: 43.73% Conservative: 64  
 Best Local Similarity: 25.07% Mismatches: 125  
 Query Match: 7.73% Indels: 68  
 DB: 1 Gaps: 13

US-09-868-987-1 (1-1864) x SECF\_STRCO (1-373)

Qy 788 ATAAAGCATGATTTCTTGAGAGGATGCAAAAACCTTTGGGCTGTTCTCGAAGTGTTTT 847  
 Db 16 ValGlyTyrAspPheValLysAsn---ArgLysIleTrp---TyrGlyIleSerIleLeu 33  
 Qy 848 CTTTAGTGTTCGCTCTCGGGTGGAGCCCTGGAAATCCGTTTGGGAATGGATTTT 907  
 Db 34 IleThrIleThrAlaIleValGlyLeuAlaValArgGlyLeuHisMetGlyIleGluPhe 53  
 Qy 908 AAAGGAGGTATGCTTTTACCTTTTAAATCAAAAGAGCATGCGATCAGC----- 955  
 Db 54 GluGlyGlyAlaValPheThr---ThrProLysAsnMetSerAlaSerValAlaGlnThr 72  
 Qy 956 -----GATGTTGCTCAAAATCGGTGGCAAACTTGTGCATAACTACAGGAAGCT 1003  
 Db 73 GluThrTrpAlaGluAlaSerGlyHisAspAlaIleValGlnLysLeuGlyAspGly 92  
 Qy 1004 GGTCTTCTCTAGAGACTTCCGTATTCAACACATTTTCGATCT-----TCAGAA 1051  
 Db 93 SerLeu-----ArgIleGlnIleAlaGlyThrAspThrGlnGlnSerAsp 107  
 Qy 1052 AAGATCAAAATCTATTTAGTATAAAGCTTTTAACTACTACTAGCAGATACGAGCTCT 1111  
 Db 108 GlnIleLysGluAspLeuSerLysAsnLeuAspValSerAlaGluLysIleAsnAlaAsp 127  
 Qy 1112 CTCTAAATTAACGATCATGAGCTGCGGTATTGTGCGGATTGTTGTCAGAAACAGGCCT 1171  
 Db 128 LeuValGly-----ProSerTrp----- 133  
 Qy 1172 AGATTCTCTACGGAAACTCTAAACGAAACGCAAAATTTTGGTCAAAAGGTAAAGCAGCAA 1231  
 Db 134 -----GlyAspGlnIleAlaAsn---LysAlaTrpGln----- 143  
 Qy 1232 CTATCGAAGAAATGCGTTATATAGCGCACCATCGGGCTTTTAGGAGCTTTGGCAATCATC 1291  
 Db 144 -----GlyLeuGlyIlePheMetValLeuVal 152  
 Qy 1292 TTGCTCTATGAGTTTGGCGTTTGAATGGCAATATCTTTCAGTCGCGTATGCGCTTTA 1351  
 Db 153 ValIleTyrLeuAlaIleAlaPheGluTrpArgMetAlaLeuAlaPheValAlaLeu 172  
 Qy 1352 ATTCATGACCTTTTGGCTACTGTCGAGCTGTGTTTATAGCACATTTCTTTTGAAGAAA 1411  
 Db 173 IleHisAspIleThrIleThrValGlyIleTyrAlaLeuValGlyPhe----- 188  
 Qy 1412 ATTCAAATAGATTTGCAAGCCATTTGCTTTAATGACTGTATTGGGGTATTCATTAAAC 1471  
 Db 189 --GluValThrProGlyThrValIleGlyLeuLeuThrIleLeuGlyTyrSerLeuTyr 207  
 Qy 1472 AATACTTTGATCATTTTGTGATCGTATTCGTGAA-----GATCCGCAAGCAACCTG 1522  
 Db 208 AspThrValValValPheAspSerLeuLysGluGlnThrArgAspIleThrLysGlnThr 227  
 Qy 1523 TTTACCCCTATGCATGTTTGTAGTTAATGATGCGCTTCAAAAGACGTTTCAGCGCACGGTA 1582  
 Db 228 ArgTrpThrTyrAlaGluIleAlaAsnArgSerIleAsnSerThrLeuValArgSerIle 247  
 Qy 1583 ATGACAACAGCTACAACCTCATCATGTTTGTGTTAATGCTTTTGTATAGGCGGC----- 1636  
 Db 248 AsnThrThrValValAlaLeuLeuProValAlaGlyLeuLeuPheIleGlyGlyVal 267  
 Qy 1637 -----TCCTCTGCTTTTAAATTTTGCATTTTATATGACCATAGGATTCCTCTAGGA 1687  
 Db 268 LeuGlyAlaGlyMetLeuAsnAspIleSerLeuSerLeuPhePheValGlyLeuAlaGly 287

QY 1688 ACTTATGCTCTTATATGACACCTCTGTTGTTATGCTCGTAAGAAAT 1747  
 Db 288 AlatyrserserilelleaThrProleuValAlaAspLeuLysGluAlaGluPro 307  
 QY 1748 CGCTCAAA 1756  
 Db 308 GlmetylS 310

RESULT 29  
 SECF\_MYCLE STANDARD; PRT; 471 AA.

ID P38386; 069486; 01-OCT-1994 (Rel. 30, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein-export membrane protein secf.  
 GN SECF OR M0448 OR MLCB1259.06 OR B1177\_C3\_239.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 ON NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smith D.R., Robison K.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=2112873; PubMed=11234002;  
 RA Cole S.T., Eiglnier K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrett B.G.;  
 RT "Massive gene decay in the leprosy bacillus";  
 RL Nature 409:1007-1011(2001).  
 CC - FUNCTION: INVOLVED IN PROTEIN EXPORT.  
 CC - SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS  
 CC WHICH COMPRISES SECA, SECB, SECD, SECE, SECF, SECG AND SECH  
 CC (BY SIMILARITY).  
 CC - SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC - SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC -----  
 CC EMBL: U00011; AAA17100.1; ALT FRAME.  
 CC EMBL: AL023591; CAA19081.1; -  
 CC EMBL: AL583918; CAC29996.1; -  
 CC Leptoma; M04488; -  
 CC InterPro: IPR003335; SecD\_SecF.  
 CC Pfam: PF02355; SecD\_SecF; 1.  
 CC TIGRFAMs: TIGR00916; 2A0604S01; 1.  
 CC TIGRFAMs: TIGR00966; 3A0501S07; 1.  
 CC Protein transport; Translocation; Transmembrane; Membrane;  
 KW Complete proteome.  
 FT TRANSMEM 88 108 POTENTIAL.  
 FT TRANSMEM 116 136 POTENTIAL.  
 FT TRANSMEM 211 231 POTENTIAL.  
 FT TRANSMEM 267 287 POTENTIAL.  
 FT TRANSMEM 325 345 POTENTIAL.  
 FT TRANSMEM 355 375 POTENTIAL.  
 SQ SEQUENCE 471 AA; 50182 MW; 1AD6D2B297A31217 CRC64;

Alignment Scores:  
 Pred. No.: 1,24e-12 Length: 471  
 Score: 252.00 Matches: 81  
 Percent Similarity: 44.55% Conservative: 66  
 Best Local Similarity: 24.55% Mismatches: 125  
 Query Match: 7.61% Indels: 58  
 DB: 1 Gaps: 10

US-09-868-987-1 (1-1864) x SECF\_MYCLE (1-471)

QY 809 GGATGCAAAAACCTTGG---GCGTTTCTGCA-----AGTGTTCCTTTTNGGT 856  
 Db 79 GlyArgArgArgLeuThrPyrSerValGlyValIleValAlaValIleValSer 98  
 QY 857 TGGCTGCTCTCGGGTTTGGAGCCTGGAAATCCGTTTGGAAATGATTTTAAAGAGG 916  
 Db 99 IleIleValArgGlyPhe-----ThrPheGlyIleAspPheLysGlyGly 113  
 QY 917 TATGCTTTCCTTATCCAAAAGAGATGCGATGCGATGCTTGTCAATGCGTGGC 976  
 Db 114 ThrThrValSerMet---ProValSerProGlyValGlyGlyThrGlyAlaIleGluVal 132  
 QY 977 AAAGTTGTCATAAATACAGAGAGAGTGTCTTCTTCTAGAGACTTCGGATTCACACA 1036  
 Db 133 AlaGlnValAlaAspValPheLysLysThrLeuGlySerAspProGluSerValValVal 152  
 QY 1037 TTTGATCTTCAGAAAAGATCAAAATCTATTTAGTATTAAGCTTTAAGC-----TAT 1090  
 Db 153 ValGlyLysGlyAlaSerValIleValThrValArgIleSerSerLysThrLeuSerAsnAspGln 172  
 QY 1091 ACTPAGAGATAGAGAGCTCTCTCTTAATTAACGATCATGAGCTGGCTTATTTGGG 1150  
 Db 173 ThrSerLysLeuArgAsnAlaLeuPheAspAlaPhe----- 184  
 QY 1151 ATTGTTGTCAAAAACAGCGCTAGATTCTCTACAGAAATCTTAACGAAACGCAAAATT 1210  
 Db 185 -----GlyProLysGlyAlaAspAlaLysProSerLysGlnAlaIleSer 199  
 QY 1211 TGGTCAAGGTAAAGACGAAACATATCGAAGAAATGCGTTATCAGGACATCGGGCTT 1270  
 Db 200 AspAlaIleValIleSerGluThrTrpGlyGlyGlnIleThrLysLysValIleAlaLeu 219  
 QY 1271 TTAGAGACTTGGCAATCATCTGCTTATGTAGTGGCTTGTGAATGACATATGCT 1330  
 Db 220 ValValPheLeuValLeuValIleGlyLeuLysIleThrValArgLysGlyLysMetAla 239  
 QY 1331 TTCATGCCGATGCGCTTAAATTCATGACCTTTTGGCTTACGTCGCTGATGCTTTTAT 1390  
 Db 240 IleSerAlaLeuThrThrMetCysPheAspPheLeuThrValThrAlaGlyValIleSerLeu 259  
 QY 1391 GCACATTTCTTTTGAAGAAATTCAAATAGATTGGACCATTTGGTGGCTTATGACT 1450  
 Db 260 ValGlyLys-----GluValThrProAlaThrValIleGlyLeuLeuThr 274  
 QY 1451 GTATGGGGTATTCATTAACAATCACTTATGATCTTTTATGATCGTATTCGTAAGAT 1507  
 Db 275 IleLeuGlyPheSerLeuLysThrValIleValPheAspLysValIleGluAsnThr 294  
 QY 1508 -----CGCCAAAGGAACCTGTTAACCCCT 1511  
 Db 295 HisGlyPheGlnHisThrThrArgArgThrPheAlaGluGlnAlaAsnLeu----- 311  
 QY 1532 ATGATGTTTATAGTATGATGCGCTTCAAAAGACGTTACGCGACGATATGACACA 1591  
 Db 312 -----AlaIleAsnGlnThrPheMetArgSerIleAsnThrSer 324  
 QY 1592 GCTAACACTATCAGTTTGTATGCTTTG-----TTTATAGCGCGC 1636  
 Db 325 LeuIleSerValLeuProValIleValAlaLeuMetValValAlaValIleLeuGlyVal 344  
 QY 1637 TCCTGCTCTTATATTTGATTTATATGACCATAGAGATTCCTTACGAACTTATGCG 1696  
 Db 1637 ----- 1696

Db	185	AlaAspGlyLysProHisProGlnGlnValIleThrValGlyAlaTyProProThrLys	204
Qy	170	CGTGAAGTGAAGCAA-----CTCGCCTCAGATTTAAATCTCGAGCGATGCTCTTT	230
Db	205	GlulIleAspGluAlaMetAlaIleTySerAlaLeuLysSerGlyAlaLeuProVal	224
Qy	221	GTTCGCCGAGTTCACAGTGAAGAGACGATCTCTTCTGATCTTCGGGAAAAACAATGTACA	280
Db	225	LysLeuAspIleGluTyIleSerThrIleSerProGluPheGlyLysGluPheLeuLys	244
Qy	281	CAGGCCATTATCTCAGCATCTGTGGCTGGCAATGCTTATGCTTTGATGACCGGTATAT	340
Db	245	GlyThrAlaIleAlaLeuLeuLeuAlaPheIleAlaValGlyIleIleValSerIleArg	264
Qy	341	TATAGATTTCGAGCGGTCATCGCTCGGGAGCTGTT-----CTTCTGAATCTTTTGCTT	394
Db	265	TyrLysGlnProLysIleAlaIleProIleLeuIleThrCysIleSerGluValIleIle	284
Qy	395	ATCTGGGCAGCTCTACAGTATTTGGATGCGCCACTCACCTTGTCAGGACTCGCTGGGATT	454
Db	285	IleLeuGlyPheAlaSerLeuIleAspTrpLysLeuAspLeuProSerIleAlaGlyIle	304
Qy	455	GTTCCTGCTATGGGATGGCGGTAGATGCAAATGTTCTTGTTATTCGAAGAATCCGAGAG	514
Db	305	IleAlaAlaValGlyThrGlyValAspAsnGlnIleValIleThrAsp-----	320
Qy	515	GAATTTTATTGTCCTCAAGCTCTTAAAAAATCTGTAGAAAA-----GGATATACC	565
Db	321	-----GluAlaLeuLysArgGlyAlaGlyLysIleArgAlaSerIleLys	335
Qy	566	AAGCCTTTTCGAGCCATTTTGTGATTTCTAACTTGACTACAGTATTGCGCTCAGCACTTCTT	625
Db	336	ArgAlaPhePheIleIlePheAlaSerAlaAlaThrSerIleAlaAlaMetLeuProLeu	355
Qy	626	TTCCTCTCAGATCAGGGCTCTTAAAGGGTTTGCTTTTGACATGTATTTAGGAATTTTC	685
Db	356	PheValLeuGlyValGlyMetLeuLysGlyPheAlaIleThrThrIleAlaGlyValLeu	375
Qy	686	TCITCAATGTTT 697	
Db	376	IleGlyIlePhe 379	
RESULT	31		
Y849	METH		
ID	Y849	METH	STANDARD; PRT; 403 AA.
AC	O26937;		
DT	30-MAY-2000	(Rel. 39, Created)	
DT	30-MAY-2000	(Rel. 39, Last sequence update)	
DE	16-OCT-2001	(Rel. 40, Last annotation update)	
DE	Hypothetical protein MTH849.		
GN	MTH849.		
OS	Methanobacterium thermoautotrophicum.		
OC	Archaea; Euryarchaeota; Methanobacteria;		
OC	Methanobacteriaceae; Methanothermobacter.		
NCBI	TaxID=187420;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Delta H;		
RC	MEDLINE=98037514; PubMed=9371463;		
RA	Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,		
RA	Alldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,		
RA	Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,		
RA	Spadafora R., Viare R., Wang Y., Wierzbowski J., Gibson R.,		
RA	Ulwani N., Caruso A., Bush D., Sater H., Patwell D., Prabhakar S.,		
RA	McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,		
RA	Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;		
RT	"Complete genome sequence of Methanobacterium thermoautotrophicum		
RT	deltaH: functional analysis and comparative genomics.";		
RL	J. Bacteriol. 179:7135-7155(1997).		
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).		
CC	-1- SIMILARITY: BELONGS TO THE SEC2/SEC7 FAMILY.		







QY 119 TTAACGCTCCATGAAATATCATGCCAGTCTCTCAGGGAATTTACCCACCGTGAAGTG 178  
Db 111 LeuAsnValPheAsnLysGluAsnTyrGlnPheSerSerThrThrLysLeuVal 30  
QY 179 AGCAAACTCGCCTCAGATTTAAATCTGGAGCGATGCTTTTGTCTCCGAGGTTCTCAGT 238  
Db 31 SerAsnSerThrValAspAlaAspAsnGlyAlaSerAspPhe-----GluAlaGlyGln 48  
QY 239 GAAGAGAGCATCTCTCTGATCTTTGGGAAACAAATGTCACACAGGCATTATCTCAGCA 298  
Db 49 GlnPheAlaThrGluLeuAspGlnGlyGlyLysGlnLeu-----GlyLeuSer--- 65  
QY 299 TGCTGTGGCTGGCAATGCTTATTTGTTGATGAGCGTATATATAGATTGGAGGCGTC 358  
Db 66 CysIleGlyLeuIleCysAsnArgMetLeuGlyThr-----GlyVal 79  
QY 359 ATCGCTTCGGGAGCTGTTCTCTCAATCTT-----TTGCTTATC 397  
Db 80 PheAlaValSerSerThrIleTyrThrLeuCysGlySerValGlyLeuAlaLeuIleMet 99  
QY 398 TGGCAGCTCTACAGTATTTGGATGCGCCACTCACCTTGTGAGGACTCGCTGGGATGTT 457  
Db 100 TrpAla-----ValGlyAlaIleIleAlaIleSerGlyLeuTyr---ValTyr 114  
QY 458 CTGCTATGGGATGGCGGTAGATGCAAAATGTTCTGTATTTCGAAGAATCCGAGAGGAA 517  
Db 115 MetGluPheGlyThrAlaIlePro----- 122  
QY 518 TTTTATTGCTCAAGCTTAAATAATCTGTAGAAAAGATATACCAAGCTTTTGA 577  
Db 123 -----LysAsnGlyGlyGlyLysAsnTyrLeuGlu----- 132  
QY 578 GCCATTTT-----GATTCTAACTGACTACAGTATTGGCTCAGCACTTCTTTTCTTC 631  
Db 133 AlaIlePheArgLysProLysPhePheIleThrCysMetTyrAlaAlaTyrIlePhePhe 152  
QY 632 CTAGATCAGGCGCTATTAAAGGTTGCTTTGATGATGATTAGGAATTTTCTTCA 691  
Db 153 Leu-----GlyTrpAlaAlaGlyAsnSerIleAsnThrAlaIleMetPheLeuThrAla 170  
QY 692 ATGTTTACGGCTCTTTCTGACTCAATTT----- 721  
Db 171 AlaAspThrGlu-----ValThrLysTrpAsnGlnArgGlyIleGlyValAlaVal 188  
QY 722 TTCCTTCATGCTGTGGATGAATAAGACCAACATACACAGTTGCATATGATGAATAAGTTC 781  
Db 189 PhePheAlaPheLeuIleAsnSer-----LeuAsnValLys 200  
QY 782 GTGGGATAAAGCATGATTTCTTGAGAGGATGCAAAAACCTTTGGCTGTTTCTGGAAGT 841  
Db 201 IleGlyLeu-----TyrLeuGlnAsn----- 207  
QY 842 GTTTTCTTTTAGCTTGGTGTCTCGGTTTGGAGCTGGAATTCGTTTGGGAATG 901  
Db 208 -----IleLeuGlyIlePheLysIleGlyIleValLeuPheIleSerIleThrGlyTrp 225  
QY 902 GATTTTAAAGAGGCTATGCTTACCTTTAACTTAAACAAAGAGCATGCGATCAGCGATGTT 961  
Db 226 ValAlaLeuGlyGly-----GlyLeuLysAsp--- 234  
QY 962 GCTCAATGCGTGCAAGTTGTGCAATAAACTACAGGAAGCTGCTTCTTCTTAGAGAC 1021  
Db 235 -----GlyTyrGlnSerIleAsn 240  
QY 1022 TTCGCTATTCAACATTTGGATCTTCAAAAAGATCAAAATCTATTTTAGTGATAAGCT 1081  
Db 241 PheArg---AsnAlaPheGluGlyThrGluThrAlaThrAlaTyrGlyIleValAsnAla 259  
QY 1082 TTAAGCTATACTAAGCAGATACGAGCTCTCTCTCTTAAATTAACGATCATGAGCTGGCGT 1141  
Db 260 Leu---TyrSer-----ValIleTrpSer 266

QY 1142 TATTGTGGGATTGTTGTCAAGAAACAGCGCCTAGATTCTCTACGGAACACTCTAAACGAAAC 1201  
Db 267 PheValGly----- 269  
QY 1202 GCAAAATTTTGGTCAAAAGTAAGCAGCAAAACTCGAAGAAATAATCGCTTATCAGCGCACC 1261  
Db 270 -----TyrSerAsnValAsnTyrAlaLeuGlyGluValLysAsnProValArgThr 286  
QY 1262 ATCGGGCTTTTAGGAGCT-----TTGGCAATCATCTTGTCTCTATGTGAGT 1306  
Db 287 LeuLysIleAlaGlyProThrSerMetValPheLeuAlaIleIleTyrIlePheValAsn 306  
QY 1307 TTGGCTTTTGAATGGCAATATGCTTTCAGTCCGATGCGCTTTAATTCATGACCTTTTG 1366  
Db 307 Ile-----AlaTyrPheAlaVal-----ValProLysAspLysLeu 318  
QY 1367 GCTACCTGTGCGAGTCTTGTATTATAGCACATTTCTTTTGAAGAAATTCAAATAGATTG 1426  
Db 319 IleSerSerLysLeuIleLeuAlaAlaAspPhe-----AspIleValPheGlyGly 336  
QY 1427 CAAGCCATTTGGTCTTAATGACTGTATTGGGGTATTCA---TTAAACAATACTTTG--- 1480  
Db 337 GlnAlaLysArgAlaAlaAlaLeuValGlyLeuSerAlaLeuGlyAsnValLeuSer 356  
QY 1481 ATCATTTTTCATCGTATTCGTGAAGATCGCAA----- 1513  
Db 357 ValIlePheSerGlnGlyArgIleIleGlnGlnLeuGlyArgGluGlyValLeuProPhe 376  
QY 1514 GCGNACTGTTTACCCCTATGTCATGTTTATGTTAAATGATGCCCTTCAAAAGACGTTTCAGC 1573  
Db 377 SerAsnPhePheAlaSerSerLysProPhe---AsnSerProMetValGlyLeuPheGln 395  
QY 1574 CGCACGGTAATGACACAGCTACAACTCTATCA-----GTTTGTG 1612  
Db 396 HisPheIleValCysThrValThrIleLeuAlaProProGlyAspAlaTyrLeuLeu 415  
QY 1613 TTAATGCTTTTGTATTATGAGCGGCTCTCTGCTTTAATTTTGCATTTATATGACCAT 1672  
Db 416 ValGlnAsnLeuIleSerTyrProMetAsnIleAlaAsnPheAla-----IleSerAla 433  
QY 1673 GGGATTCCT----- 1681  
Db 434 GlyLeuLeuTrpIleTyrTrpGlnArgGlnGlyLysIleGluTrpAsnProProIle 453  
QY 1682 -----CTAGGAACCTTATCGTCTCTTTAT-----ATTGCA 1711  
Db 454 LysAlaGlyValPheValThrGlyPhePheThrLeuSerAsnLeuTyrLeuIleAla 473  
QY 1712 CCA 1714  
Db 474 Pro 474  
RESULT 33  
ACRF\_ECOLI  
ID ACRF\_ECOLI STANDARD; PRT; 1034 AA.  
AC P24181;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-OCT-2001 (Rel. 40, Last annotation update)  
DE Acriflavine resistance protein F (Protein envd).  
GN ACRF OR ENV0 OR B3266.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia  
OC NCBI\_TaxID=562;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=K12;  
RA Xu J., Nilles M.L., Bertrand K.P.;  
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP PRELIMINARY SEQUENCE FROM N.A.  
RC STRAIN=K12;

RX MEDLINE=92079901; PubMed=1720861;  
 RA Klein J.R., Hentrich B., Piapp R.;  
 RT "Molecular analysis and nucleotide sequence of the envC operon of  
 RT Escherichia coli.";  
 RL Mol. Gen. Genet. 230:240-240(1991).  
 RN [3]  
 RP REVISIONS.  
 RX MEDLINE=94012493; PubMed=8407802;  
 RA Ma D., Cook D.N., Alberti M., Pon N.G., Nikola H., Hearst J.E.;  
 RT "Molecular cloning and characterization of acra and acrf genes of  
 RT Escherichia coli.";  
 RL J. Bacteriol. 175:6299-6313(1993).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1232-1274(1997).  
 CC -1- FUNCTION: INVOLVED IN CELL ENVELOPE FORMATION. IS PRODUCED IN  
 CC EXTREMELY LOW AMOUNTS.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
 CC -1- SIMILARITY: BELONGS TO THE ACRB/ACRD/ACRF FAMILY.  
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO  
 CC FRAMESHIFTS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; M96848; AAA02932.1; -;  
 DR EMBL; X57948; CAA41017.1; -;  
 DR EMBL; U18997; AAB58070.1; -;  
 DR EMBL; AE000405; AAC76298.1; -;  
 DR PIR; S18537; S18537.  
 DR Ecogene; EG10267; acrf.  
 DR InterPro; IPR001036; ACRF\_vin\_res.  
 DR InterPro; IPR004764; HAE1.  
 DR Pfam; PF00873; ACRF\_tran\_1.  
 DR PRINTS; PR00702; ACRIFLAVINRP.  
 DR TIGRFAMs; TIGR00915; 2K0602; 1.  
 KW Cell division; Transmembrane; Inner membrane; Transport;  
 KW Complete proteome.  
 FT DOMAIN 1 9  
 FT TRANSMEM 1 28  
 FT DOMAIN 29 339  
 FT TRANSMEM 340 359  
 FT DOMAIN 360 365  
 FT TRANSMEM 366 385  
 FT DOMAIN 386 391  
 FT TRANSMEM 392 413  
 FT DOMAIN 414 441  
 FT TRANSMEM 442 460  
 FT DOMAIN 461 473  
 FT TRANSMEM 474 496  
 FT DOMAIN 497 537  
 FT TRANSMEM 538 556  
 FT DOMAIN 557 871  
 FT TRANSMEM 872 891  
 FT DOMAIN 892 897  
 FT TRANSMEM 898 917  
 FT DOMAIN 918 923  
 FT TRANSMEM 924 945  
 FT DOMAIN 946 973  
 FT TRANSMEM 974 992  
 FT DOMAIN 993 1005

FT TRANSMEM 1006 1028 12 (BY SIMILARITY).  
 FT DOMAIN 1029 1034 CYTOPLASMIC (BY SIMILARITY).  
 SQ SEQUENCE 1034 AA; 111454 MW; B676807AE5BD17B1 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.00689 Length: 1034  
 Score: 144.00 Matches: 135  
 Percent Similarity: 31.73% Conservative: 89  
 Best Local Similarity: 19.12% Mismatches: 226  
 Query Match: 4.35% Indels: 256  
 DB: Gaps: 30  
 US-09-868-987-1 (1-1864) x ACRF\_ECOLI (1-1034)  
 QY 308 TTGGCAATGCTTATGTTGATGAGCGTATATTATGATTGGAGCGCATCGCTTCG 367  
 DB 353 LeuValMetTyrLeuPheLeuGlnAsnMetCysAlaThrLeuIleProThrIleAlaVal 372  
 QY 368 GAGCGTGTCTGATCTTTGCTTATCTGGCGAGCTCTACAGATTGTGGATGCCCA 427  
 DB 373 ProValValLeuLeuGlyThrPheAlaIleLeuAlaIleGlyTyrSerIleAsnThr 392  
 QY 428 CTCACCTTGTACAGACTCGCTGGGATTGTTCTTCTATGGGATGGCCGTAGATGCAAT 487  
 DB 393 LeuThrMetPheGlyMet-----ValLeuAlaIleGlyLeuLeuValAspAla 409  
 QY 488 GTTCTGTATTGCAAGATCCAGAGAAATTTTATTTGTCGTAAGCTTT-----AAA 541  
 DB 410 IleValValValGlnAsnVal---GluArgValMetMetGlnAspTyrLeuProPhe 428  
 QY 542 AAATCTGATGAAAGATATACCAAGCGCTTTGGAGCGCATTTTGAATCTTAATCGACT 601  
 DB 429 GlnAlaThrGlnLeuSerMetCysGlnIleGlnAlaLeuVal-----GlyIleAla 446  
 QY 602 ACAATATGGCTCAGACACTTCTT-----TTCTCTCATGATACAGGCGCTATT 649  
 DB 447 MetValLeuSerAlaValPheIleProMetAlaPhePheGlySerThrGlyAlaIle 466  
 QY 650 ---AAAGGTTTGCTTGAACATGATTATAGAAATTTCTTCATGATTAGCGCTCTT 706  
 DB 467 TyrArgGlnPheSerIleThrIleValSerAlaMetAlaLeuSerValLeuValAlaLeu 486  
 QY 707 TTCAATGACT----- 715  
 DB 487 IleLeuThrProAlaLeuCysAlaThrLeuLeuSerProValSerAlaGlnHisGln 506  
 QY 716 ---AAATTTTCTTCATGCTGTGGATGATAGAAC---CAATACACAGTTGCATATG 769  
 DB 507 AsnLeuGlyGlyPhePheGlyTyrPheAsnThrThrPheAspHisSerValAsnHisTyr 526  
 QY 770 ATGAAT-----AAGTTGTTGGGATTAAGCATGATTTCTTG----- 805  
 DB 527 ThrAsnSerValGlyIleLeuGlySerThrGlyArgTyrLeuLeuIleTyrAlaLeu 546  
 QY 806 ---AGAGATGCCAAAAAATTTGGCGCTGTTCTGAGATGTTTCTT----- 850  
 DB 547 IleValIleGlyMetValValLeuPheLeuArgLeuProSerSerPheLeuProGlnGlu 566  
 QY 851 ---TTAGTTGGCTGCTCTCGGCTTGGAGCTTGAGAT----- 886  
 DB 567 AspGlnIleValPheLeuThrMetIleGlnLeuProAlaGlyAlaThrGlnGluArgThr 586  
 QY 886 ----- 886  
 DB 587 GlnIysValLeuAspGlnValThrAspTyrTyrLeuIysAsnGlnIysAlaAsnValGlu 606  
 QY 887 -----TCGTTTGGGAATGATTTTAAAGA----- 913  
 DB 607 SerValPheThrValAsnGlyPheSerPheSerGlyGlnIleAsnAlaGlyMetAla 626  
 QY 913 ----- 913  
 DB 627 PheValSerLeuIysProTyrGlnGluArgAsnGlyAspGlnAsnSerAlaGluAlaVal 646

QY 914 -----GGGTATGCGCTTTACCTTTAAAT 934  
Db 647 ILeHisArgAlaLysMetGluLeuGlyLysIleArgAspGlyPheValIleProPheAsn 666  
QY 934 ----- 934  
Db 667 MetProAlaIleValGluLeuGlyThrAlaThrGlyPheAspPheGluLeuIleAspGln 686  
QY 934 ----- 934  
Db 687 AlaGlyLeuGlyHisAspAlaLeuThrGlnAlaArgAsnGlnLeuLeuGlyMetAlaAla 706  
QY 935 -----CCAAAGAGCATGCATCAGCGATGTGTGCTCAATG 970  
Db 707 GlnHisProAlaSerLeuValSerValArgProAsnGlyLeuGluAspThrAlaGlnPhe 726  
QY 971 CGTGGCAAGTT---GTGCATAAATCAGAGAGCTGTCTTCTTCTAGAGACTTCGT 1027  
Db 727 LysLeuGluValAspGlnGluLysAlaGlnAlaLeuGlyValSerLeuSerAspIleAsn 746  
QY 1028 ATTCAAACATTTGGATCTCTCA----- 1048  
Db 747 ---GlnThrIleSerThrAlaLeuGlyThrTyrrValAsnAspPheIleAspArgGly 765  
QY 1049 -----GAAAGATCAAAATCTATTTAGTGATATAA 1078  
Db 766 ArgValLysLysLeuTyrrValGlnAlaAspAlaLysPheArgMetLeuProGluAspVal 785  
QY 1079 GCTTTAAGCTATCTAGCAGATACGACCTCTCTCTA-----AAATTACGATC 1129  
Db 786 AspLysLeuTyrrValArgSerAlaAsnGlyGluMetValProPheSerAlaPheThrThr 805  
QY 1130 ATGAGCTGG-----CGTTATTGTGGGATT---GTTGTGAGA 1162  
Db 806 SerHisIrrValTyrrGlySerProArgLeuGluArgTyrrAsnGlyLeuProSerMetGlu 825  
QY 1163 AACAGGCTAGATTCTTACGAACTCTAAACGAACGCA---AAATTGTGTCAAAG 1219  
Db 826 ILeGlnGlyGluAlaAlaProGlyThrSerSerGlyAspAlaMetAlaLeuMetGluAsn 845  
QY 1220 GTAAGCAGCAACTATCG-----AGAAATCGCTTATCAGCG 1258  
Db 846 LeuAlaSerLysLeuProAlaGlyIleGlyTyrrAspTrpThrGlyMetSerTyrrGlnGlu 865  
QY 1259 ACCATCGGG-----CTTTAGAGCTTTGGCAATCATCTGCTCTATGTG 1303  
Db 866 ArgLeuSerGlyAsnGlnAlaProAlaLeuValAlaIleSerPheValValIlePheLeu 885  
QY 1304 AGTTTGGCTTTGAATGCAATATGCTTTTCAGTCCGATCGCTTTAAATTCATGACCTT 1363  
Db 886 CysLeuAlaAlaLeuTyrrGluSerTrpSerIleProValSerValMetLeuValValPro 905  
QY 1364 TTGCTACCTGTGCAGCTCTCTTTATAGCATTTCTTTTGAAGAAATTCAAATAGAT 1423  
Db 906 LeuGlyIleValGlyValLeuLeuAlaThrLeuPheAsnGlnLys-----AsnAsp 923  
QY 1424 TTGCAAGCATTGTGCTTTAATGACTGTATTTGGGTATTCATTAACAATCTTTGATC 1483  
Db 924 ValTyrrPheMetValGlyLeuLeuThrIleGlyLeuSerAlaLysAsnAlaIleLeu 943  
QY 1484 ATTTTTCATCTATTTCGGAAGATGCGCAAGCAACCTGTTTACCCCTATGATGTTT 1543  
Db 944 ILeValGlu-----Phe 947  
QY 1544 GTTAATGATGCTTCAAAAG-----ACGTTACGCCGACGGTAATGACACA 1591  
Db 948 AlaLysAspLeuMetGluLysGluGlyValValGluAlaThrLeuMetAlaVal 967  
QY 1592 GCTCAACTCTCATGATTTGCTTAATG-----CTTTGTTTATA----- 1630  
Db 968 ArgMetArgLeuArgProIleLeuMetThrSerLeuAlaPheIleLeuGlyValLeuPro 987

QY 1631 -----GGCGGCTCTCTGCTCTTTAAATTTTGCATTTATATATGACCATA 1672  
Db 988 LeuAlaIleSerAsnGlyAlaGlySerGlyAlaGlnAsn-----AlaValGlyIle 1004  
QY 1673 GGGATTCTCTAGAACTTTATCGTCTCTTTATATTCGA-----CCACCTCTGTG 1723  
Db 1005 GlyValMetGlyMetValSerAlaThrLeuLeuAlaIlePhePheValProValPhe 1024  
QY 1724 TTGTTTATGTCGCTAA 1741  
Db 1025 PheValValIleArgArg 1030  
RESULT 34  
Y895 HAEIN STANDARD; PRT; 1032 AA.  
AC Q57124; O05034;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein HI0895.  
GN HI0895.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA McLennan A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.B., Frichman J.L., Fuhrmann J.U., Geoghagen N.S.M.,  
RA Gnehm C.B., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
Rd.";  
RL Science 269:496-512 (1995).  
CC -!- FUNCTION: COULD BE A DRUG EFFLUX PUMP (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
(By similarity).  
CC -!- SIMILARITY: BELONGS TO THE ACRB/ACRD/ACRF FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U32771; AAC22555.1; -;  
DR TIGR; HI0895; -;  
DR InterPro; IPR001036; Acrflvin\_res.  
DR Pfam; PF00873; ACR tran; 1.  
DR PRINTS; PR00702; ACRFLVINRP.  
KW Hypothetical protein; Transmembrane; Inner membrane; Transport;  
KW Complete proteome.  
FT DOMAIN 1 17 CYTOPLASMIC (BY SIMILARITY).  
FT TRANSMEM 18 36 1 (BY SIMILARITY).  
FT DOMAIN 37 337 PERIPLASMIC (BY SIMILARITY).  
FT TRANSMEM 338 357 2 (BY SIMILARITY).  
FT DOMAIN 358 363 CYTOPLASMIC (BY SIMILARITY).  
FT TRANSMEM 364 383 3 (BY SIMILARITY).  
FT DOMAIN 384 389 PERIPLASMIC (BY SIMILARITY).  
FT TRANSMEM 390 411 4 (BY SIMILARITY).  
FT DOMAIN 412 438 CYTOPLASMIC (BY SIMILARITY).  
FT TRANSMEM 439 457 5 (BY SIMILARITY).  
FT DOMAIN 458 470 PERIPLASMIC (BY SIMILARITY).  
FT TRANSMEM 471 493 6 (BY SIMILARITY).

FT DOMAIN 494 529 CYTOPLASMIC (BY SIMILARITY).  
 FT TRANSMEM 530 548 7 (BY SIMILARITY).  
 FT DOMAIN 549 852 PERIPLASMIC (BY SIMILARITY).  
 FT TRANSMEM 853 872 8 (BY SIMILARITY).  
 FT DOMAIN 873 878 CYTOPLASMIC (BY SIMILARITY).  
 FT TRANSMEM 879 898 9 (BY SIMILARITY).  
 FT DOMAIN 899 910 PERIPLASMIC (BY SIMILARITY).  
 FT TRANSMEM 911 932 10 (BY SIMILARITY).  
 FT DOMAIN 933 960 CYTOPLASMIC (BY SIMILARITY).  
 FT TRANSMEM 961 979 11 (BY SIMILARITY).  
 FT DOMAIN 980 992 PERIPLASMIC (BY SIMILARITY).  
 FT TRANSMEM 993 1015 12 (BY SIMILARITY).  
 FT DOMAIN 1016 1032 CYTOPLASMIC (BY SIMILARITY).  
 SQ SEQUENCE 1032 AA; 111633 MW; C06658844C8F3C90 CRC64;

## Alignment Scores:

Pred. No.: 0.00277 Length: 1032  
 Score: 136.50 Matches: 136  
 Percent Similarity: 34.91% Conservative: 145  
 Best Local Similarity: 16.89% Mismatches: 241  
 Query Match: 4.12% Indels: 283  
 DB: 1 Gaps: 30

US-09-868-987-1 (1-1864) x Y895\_HAEIN (1-1032)

QY 44 AATGACAAATTTCTGCAAAACCGGATGGCTATGCTAGTATGAGTTATATG 103  
 DB 274 AshnerigalathirialaanglyalaguserValValleualleashprothser 293  
 QY 104 GTGACAGCCCTATTTTAAAGTCGCATTTAAATCATGCGAGTGTCTGAGGAAATT 163  
 DB 294 ThrilaenProleu-----ThValAlGluVal--- 303  
 QY 164 ACCACCGTAGAGGAGCAAACTGCGCTCAGATTAAATCGAGCGATG--TCITTT 220  
 DB 304 -----ileargProleutyrguseriletyrthgluLeuproAspermet 319  
 QY 221 GTTCCCGAGTTCATGAGGAGAGACGATCTCTGATCTTGGGAAAAAACAATGTCA 280  
 DB 320 GluseripileutyryAspargthrilealileashenserSerilehiglValile 339  
 QY 281 CAAGCATTTATCTGAGATGCTGTGGCAATGCTTATTTGTTGATGAGCGTATAT 340  
 DB 340 LysThrile-----GlygluValthrileuileValleuValileu 354  
 QY 341 TATAGATTGAGGC-----GTCAATCGCTTGGAGCTGTCTTCTGAATCTT 388  
 DB 355 MetheileglyserPheArgAlalileuileProilleuAlalileProilleu 374  
 QY 389 TTGCTTATCTGGGAGCTCTACAGATTGGATGGCCATCACCCTGTGACGACTCGCT 448  
 DB 375 IleglyValleuMetLeuLeuInserPheasnPheSerileasnMetThrileu 394  
 QY 449 GGGATTTGTTCTGATGGGATGGCCATGATGCAATGTTCTGTATTTGAAAGATC 508  
 DB 395 AlaleuileuAlalileglyleuValValleuAsparAlalileValleuGluasnile 414  
 QY 509 CGAGAGAAATTTTATTTGCTCAAAAGCTTAAAAAATCTGAGAAAAAGATATACAG 568  
 DB 415 -----AsparghisiletyrsalaglygluThrPro 424  
 QY 569 GCTTTGGAGCGATTTTGTATCTAC----- 595  
 DB 425 PheargAlalaleileglyThrargGluilealValProValileSerMetThrile 444  
 QY 596 ---TTGATACAGATTTGGCGTACGACACTCTTTCTTCTAGATACAGGCGTATTT-- 649  
 DB 445 AlaleuilealavalTySerProMetAlaleuMetGlyglylethrGlyThrleuPhe 464  
 QY 650 AAAGGTTTGCTTGACATTTGATTTAGAAATTTTCTCTCAATGTTTACGAGCTTTTC 709  
 DB 465 LysgluPheAlaleuThrleuAlaglyAlavalPheileSerGlyValValAlaleuThr 484

QY 710 ATGACAAATTTTCTTCATG-----CTGTGATG--- 739  
 DB 485 LeuserProMetMetSerSerlySleuLeuLysSerAsnAlalyProthrTrpMetGlu 504  
 QY 740 AATAGACCCCAATACA-----CACTTG 763  
 DB 505 GluArgAlagluThrSerThrleuGlyLysValasnArgValtyrGluTyrmMetLeuAspLeu 524  
 QY 764 CATATGATGATTAAGTTCCGTGGGATTAAGCATGATTTCTTGAGAGATCAAAAACCT 823  
 DB 525 ValMetLeuasnArg-----LysSerMet 532  
 QY 824 TGAGCTTCTCTGGAAGTGTCTTTCTTTA----- 853  
 DB 533 LeuAlaPheAlavalVallePheSerThrleuProPheLeuPheasnSerLeuSerSer 552  
 QY 854 -----GGTTGCGTTCCTCGGTTTGGAGCTTGAAT 886  
 DB 553 GluLeuThrProasnGluAspLysGlyAlaPheilelealilegly---AsnAlaProSer 571  
 QY 887 TCCGTT----- 892  
 DB 572 SerValasnValAspTyrlileGlnAsnAlaMetGlnProTyrmetyAsnValMetGlu 591  
 QY 893 ---TTGGATGATTTTAAAGAGCGTATGCTTTACCTTTAATCA 937  
 DB 592 ThrProGluValSerPheGlyMetSerilealagly-----AlaProThrSerAsnSer 609  
 QY 938 AAAGAGCATGCAATCAGC-----GATGTTGCTCAATGCGTGCAGAAAGTT 982  
 DB 610 SerLeuasnleilethrThrleuLysAspTyrgluArgSerArgLysGlnSerAlaile 629  
 QY 983 GTGATAAATCTACAGAGAGCTGCTTTCTTCTAGAGACTTCGATTCAAACATTT-- 1039  
 DB 630 MetasnGluLeuasnGluLysAlalySerleProGluValSerAlaPheasn 649  
 QY 1039 ----- 1039  
 DB 650 IleProGluLeuAspThrnglyGluGlnGlyProProValSerileValleuLysThrAla 669  
 QY 1040 -----GGATCTTCAGAAAAGATC----- 1057  
 DB 670 GluAspTyryLysSerileuAlaasnThrAlagluLysPheLeuSerAlaMetLysAlaSer 689  
 QY 1058 ---AAATCTATTTTACTGATTAAGCTTATAGCTAT--ACTAGACGATACAGACTCT 1111  
 DB 690 GlyLysPheiletyrThrAsnLeuAspLeuThrTyraSphThrAlaglnMetThrileSer 709  
 QY 1112 CTGCTTAAA-----TTAACGATCATG----- 1132  
 DB 710 ValAspLysGluLysAlaglyThrTyrglyIleThrMetGlnGlnIleSerAsnThrLeu 729  
 QY 1133 -----ACGTGGCGT 1141  
 DB 730 GlySerPheLeuSerGlyAlaThrValThrArgValAspValAspGlyArgAlaTyryLys 749  
 QY 1142 TATTTGGGAGTTGTGTCAGAAC-----AGCGCTAGATT 1177  
 DB 750 ValileSerGlnValLysArgAspAspArgLeuSerProGluSerPheGlnAsnTyryr 769  
 QY 1178 CTCTACGGAATCTCTAAGCAAAAGCAAAATTTTGGCAAGGTAGACACCAAACTATCG 1237  
 DB 770 LeuThrAlaSerasnGlyInserValProLeuSerSerValileSerMetLysLeuGlu 789  
 QY 1238 AAGAAA----- 1243  
 DB 790 ThrGlnProThrSerLeuProArgPheSerGlnLeuasnSerAlagluIleSerAlaVal 809  
 QY 1244 -----ATGCGTTATCAGGCGAC-- 1261  
 DB 810 ProMetProGlyLysSerSerGlyAspAlailealatrPheGlnGlnGlnAlaThrAsp 829  
 QY 1261 ----- 1261

Db 830 AsnLeuProGlnGlyThrPheAspPheLysSerGluAlaArgGlnLeuValGlnGlu 849

Qy 1262 -----ATCGGGCTTTTGGAGCTTTGGCAATCACTTTGCTCTATGTCAGTTTGGC 1312

Db 850 GlyAsnAlaLeuAlaValThrPheAlaLeuAlaValIleIlePheLeuValLeuAla 869

Qy 1313 TTTGATGCGCAATATGCTTTTCAGTCGCGTATGCGCTTAAATTCATGACCTTTTGGCTACC 1372

Db 870 IleGlnPheGluSerIleArgAspProMetValIleMetIleSerValProLeuAlaVal 889

Qy 1373 TGTGCGAGCTTTG-----TTTATAGCACATTTCTTTTGAAGAAATTCAAATA 1420

Db 890 SerGlyAlaLeuValSerLeuAsnIleLeuSerPhePheSerIleAlaGlyThrThrLeu 909

Qy 1421 GATTTCAGCCATTGGTCTTTAATGACTGTATTGGGTATTTCATTAACAACAACTTTG 1480

Db 910 AsnIleTySerGlnValGlyLeuIleThrLeuValGlyLeuIleThrLysHisGlyIle 929

Qy 1481 ATCATTTTGTATGCTATTCGTCGATCGCCAGCGAACCTGTTTACCCCTATGATGTT 1540

Db 930 LeuMetCysGluValAlaLysGluGlnLeuAsnHisGlyLysThrArgIleGluAla 949

Qy 1541 TTAGTTAATGATGCTTCAAAAGACGTTACGCCACGCGTAATGACACACAGCTACAACT 1600

Db 950 IleThrHis---AlaAlaLysValArgLeuArgProIleLeuMetThrThrAlaAlaMet 968

Qy 1601 CTATCAGTTTGTAAATGCTTTGTTTATAGCGGCTCC-----TCGTCTTTAAAT 1651

Db 969 ValAlaGlyLeuIleProLeuLeuTyThrAlaThrGlyAlaGlyAlaValSerArgPheSer 988

Qy 1652 TTTGATTTATATGACCATGAGGATTTCTTAGAAGCTTTATGCTCTCTTTATATGCA 1711

Db 989 IleGlyIleValIleValAlaGlyLeuSerIleGlyThrIlePheThrLeuPheValLeu 1008

Qy 1712 CCA-----CCTCTGTTGTTGTTGTTTATGTCGCT 1738

Db 1009 ProValValTySerTyThrValAlaThrGluHisLysProLeuProValPhe----- 1025

Qy 1739 AAAGAAATCGCTCA 1753

Db 1026 AspGluAsnLysThr 1030

RESULT 35

MMLB STRCO

ID MMLB STRCO STANDARD; PRT; 847 AA.

AC OS4101;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

DE Putative membrane protein SC05905.

GN SC05905 OR SC10A5.10C.

OS Streptomyces coelicolor.

OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.

OX NCBI\_taxid=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2) / M145;

RX MEDLINE=21996410; PubMed=12000953;

RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,

RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,

RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,

RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howarth S.,

RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,

RA Rabinowitz S.E., Rajandream M.A., Rutherford K., Rutter S.,

RA Segger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,

RA Wier T., Wietzorrek A., Woodward J., Barrrell B.G., Parkhill J.,

RA Hopwood D.A.

RT "Complete genome sequence of the model actinomycete Streptomyces

coelicolor A3(2).";

RL Nature 417:141-147(2002).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO THE MMLB FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to license@sib-sib.ch).

CC

CC EMBL; AL021529; CAA16442.1; -

DR InterPro; IPR000731; HMGR/patch\_5TM.

DR InterPro; IPR004869; MMLB.

DR Pfam; PF031176; MMLB; 2.

DR PROSITE; PS0156; SSD; 2.

DR

KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 18 38 POTENTIAL.

FT TRANSMEM 187 207 POTENTIAL.

FT TRANSMEM 215 235 POTENTIAL.

FT TRANSMEM 248 268 POTENTIAL.

FT TRANSMEM 302 322 POTENTIAL.

FT TRANSMEM 326 346 POTENTIAL.

FT TRANSMEM 381 401 POTENTIAL.

FT TRANSMEM 539 559 POTENTIAL.

FT TRANSMEM 562 582 POTENTIAL.

FT TRANSMEM 600 620 POTENTIAL.

FT TRANSMEM 643 663 POTENTIAL.

FT TRANSMEM 672 692 POTENTIAL.

FT TRANSMEM 744 764 POTENTIAL.

FT TRANSMEM 778 798 POTENTIAL.

FT TRANSMEM 808 828 POTENTIAL.

SQ SEQUENCE 847 AA; 88616 MW; E2A443F5C74AE1DE CRC64;

Alignment Scores:

Pred. No.: 0.00358 Length: 847

Score: 135.00 Matches: 123

Percent Similarity: 33.33% Conservative: 80

Best Local Similarity: 20.20% Mismatches: 244

Query Match: 4.08% Indels: 162

DB: 1 Gaps: 23

US-09-868-987-1 (1-1864) x MMLB\_STRCO (1-847)

Qy 173 GAGTGAGCAAACTCGCTTCAGATTTAAATCTGGACGATGCTTTTGTGTT-----CCC 226

Db 155 GluValArgAlaLeuAlaSerArgAlaGlyProAlaValAlaValThrGlyPro 174

Qy 227 GAGGTTCTCAGTGAAGACGATCTCTTCGATCTTCGGGAAAAACAATGTACACAGGC 286

Db 175 AlaGlyIleAlaThrAspThrValIleValPheSerGlyGlyAspLys----- 190

Qy 287 ATTATCTCAGCATGCTGCTGCTTGGCAATGCTTATCTTTGATGAGCGTATATTATAGA 346

Db 191 ValLeuLeuAlaThrValValLeuValLeuIleLeuLeuAlaIleTyArg--- 209

Qy 347 TTTGGAGCGCTACGCTTCGGGAGCTGTTCTTCTGAATCTTTTGTATCTTCGCGGACGT 406

Db 210 -----SerProLeuMetAlaLeuValProLeuLeuAlaValGlyValAla 224

Qy 407 CTACAGTATTTGGATGCGCCACTCCTTCAGGACTCGCTGGGATT----- 454

Db 225 MetArgValAlaGluThrLeuGlyAlaIleLeuAlaAspAlaGlyValIleThrValSer 244

Qy 455 -----GTTCTTGTATGCGGATGCGCGTAGATGCAATGTT 490

Db 245 SerGlnThrAlaSerIleMetThrValLeuLeuPheGlyValGlyThrAspTyAlaLeu 264

Qy 491 CTTGATTCGAAAGAACCCGAGAGAAATTTTATTTGTTCTCAAGTCTTAAATAATCTGTA 550

Db 265 IleIleThrAlaArgTyArgGluThrLeuLeuAspGluProAspArgAlaAlaMet 284

Qy 551 GAAAGAGATATACCAAGGCTTTTGGAGCCATTTTTCATCTTAAGTCTTACACTGACATATTG 610

```

Db 285 GlnAlaIaIaValArgArgThrAlaGluSerValIleuAlaSerAlaSerThrIleValLeu 304
Qy 611 GCCTCAGACATCTTTTCTTCCTAGATACAGGGCCTATTAAAGGTTGCTTGACATGG 670
Db 305 AlamecPheAlaLeuValAlaIaValSerProAlaLeuHisGlyPheGlyProTyrLeu 324
Qy 671 ATTTAGAATTTTCTCTCAATGTTTACGGCTCTTTCATGACACAAATTTTCTTCAG 730
Db 325 AlaIeuGlyValAlaIaValMetAlaLeuValAlaPheThrPheIleProAlaLeuValLeu 344
Qy 731 CTGTGATGATGAATGAAGCCCAACATACAGTTCATATGATGAATGAATGCTGGGAGATA 790
Db 345 Leu----- 345
Qy 791 AAGCATGATTTCTTGAGAGATGCAAAAACTTGGGCTGTTTCTGGAAGTGTCTTCTT 850
Db 346 -----LeuGlyArgSerValPheTyr 352
Qy 851 TTAGTTCGCTT-----GCTCTGGGTTTGAGAGCTTGGAATTCGCTT 892
Db 353 ProGlyGlyValAspGlyAlaIaGluArgSerArgGlyAlaGlyIleTyrHisArgIle 372
Qy 893 TTGGGATGATGATTTAAAGGA----- 913
Db 373 AlaAlaLeuValAlaIaArgAlaProValIysValAlaSerAlaValIleAlaLeuVal 392
Qy 914 -----GGGTATGCCCTTACCTTTAAT----- 934
Db 393 ValIeuSerAlaGlyLeuIeuGlyTyrGlnIeuSerPheAlaThrLeuSerGlyPheArg 412
Qy 935 -----CCAAAAGACATGCGATCGATGCTGCTCAATGCCGGGAAA----- 979
Db 413 AlaAlaThrGluSerGluHisGlyGlnHisLeuIleArgGluGluPheGlyProGlyGlu 432
Qy 980 -----GTTGTCATTAACCTACAGAAAGCTGCTTCTTCTAGAGAC 1021
Db 433 IleAlaProSerThrValValValHisSerGlnAspAlaLeuArgSerSerProAlaPro 452
Qy 1022 TTCCGATTTCAACA----- 1036
Db 453 AlaAspIleAlaThrAlaLeuThrAspAlaAspHisValSerArgValAlaAspProArg 472
Qy 1037 TTTGGATCTTCAGAAAACATCAAAATCTAT-----TTTAGT 1072
Db 473 MetGlyAspGlyIleThrValPheTyrAspValIleLeuAspLeuAspProTyrSer 492
Qy 1073 GATAAGCTTTAAGCTATACAGAGATACAGAGCTCTCTCCCTAAATTAACGATCATG 1132
Db 493 SerIysAlaLeuAspAlaIle-----GlyProIeuIysGlnAlaThrGlnSer 508
Qy 1133 AGCTGCGCTTATTTGGGATTTGTTGTCAGAAAACAGGCTTAGATTTCTACGAAAACCT 1192
Db 509 AlaIaGlnAlaIaGly-----ValGlnAspAlaThrValLeuIleGlyGlyIuThr 526
Qy 1193 AAACGAAACGCAAAATTTGGTCAAGGTAAAGCAGCAACATCTGCAAGAAAATGCGTAT 1252
Db 527 AlaGlnAsnAla-----AspIleArgSerAlaLeuAspArg----- 538
Qy 1253 CAGGAGACATCGGGCTTTAGAGACTTGGCAATCATCTGCTATGATGAGTTGGCGC 1312
Db 539 AspThrThrLeuIleValLeuValLeuAlaIleValIleThrValValLeuValLeu 558
Qy 1313 TTTCGAATGCAATATGCTTTACAGTCCGTAAGCGCTTTAATTCATGACCTTTGGCTACC 1372
Db 559 LeuArgSerLeuLeuAlaIleProLeuTyrLeuValAlaIaThrLeu-----LeuIeuSerPhe 576
Qy 1373 TGTGACGCTCTGTTTATGACACATTTCTTTTGAAGAAAATTCAAATGATTTGCAACC 1432
Db 577 LeuAlaThrLeuGlyAlaIleThrPhePheThrValThrValLeuGlyAsp---AspGly 595
Qy 1433 ATTTGGT-----GCTTAATGACGTATGAGGGGATTCATTA 1468
Db 596 IleGlyAsnArgValThrAlaTyrIlePheValPheLeuValAlaIaLeuGlyValAspTyr 615

```

```

Qy 1469 AACAAATCTTGATCATTTTGTATCGATTCCGTAAGATCCCAAGCAACCTGTTTACC 1528
Db 616 Asn-----IlePheIleMetSerArgPheIleGlnIleuArgThrGln----- 630
Qy 1529 CCTATGATGATTTTATGATTAATGATGACCTTCAAAAGAG-----TTCAGCCGACG 1579
Db 631 ProProAlaIalysAlaIleIleThrAlaIaIaLeuThrArgThrIleGlyValIleSerSerAla 650
Qy 1580 GTAATGACAAACAGCTACCAACTATACGTTTGTTAATGCTTTGTTTATAGGCGGCTCC 1639
Db 651 GlyIleuIleLeuAlaIaIaThrPheAlaIaValIleuMetThrGlnProIle-----Arg 667
Qy 1640 TCTGCTCTTAATTTTGCATTTATTAATACCATATGGGATTTCTTAGAATCTTATGCTCT 1699
Db 668 GluIeuPheGlnPheGlyPheAlaMetAlaCysGlyIleLeuIeuAspThr----- 684
Qy 1700 CTTTATATGACACACCTCTGTTGTTG 1726
Db 685 ---PheIleuIleArgProIleuVal 692

RESULT 36
YC53 METUA STANDARD; PRT; 282 AA.
AC 058650;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1253.
GN MJ1253.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
ON NCBI_TaxID=2190;
RX MEDLINE=96337999; PubMed=8688087;
RA Bait C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kane B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -SIMILARITY: BELONGS TO THE SECY/SECF FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U67565; AAB99256.1; -.
CC TIGR: MJ1253; -.
CC InterPro: IPR003335; SecD_SecF.
CC Pfam: PF02355; SecD_SecF.1.
CC Hypothetical protein; Protein transport; Translocation; Transmembrane;
CC Complete proteome.
CC
CC TRANSMEM 9 29 POTENTIAL.
CC TRANSMEM 120 140 POTENTIAL.
CC TRANSMEM 149 169 POTENTIAL.
CC TRANSMEM 174 194 POTENTIAL.
CC TRANSMEM 214 234 POTENTIAL.
CC TRANSMEM 236 256 POTENTIAL.
CC SEQUENCE 282 AA; 31029 MW; D2AC5859AERC1079 CRC64;

```

Alignment Scores:			
Pred. No.:	Score:	Length:	282
Percent Similarity:	49.12%	Conservative:	42
Best Local Similarity:	24.56%	Mismatches:	51
Query Match:	3.99%	Indels:	36
DB:	1	Gaps:	7
US-09-868-987-1 (1-1864) x YCS3_METUA (1-282)			
QY	236	AGTGAAGAGACGACTCTCTCTGATCTTTGGGAAAAACAATGTATACAAAGGCATTATCTCA	295
Db	105	SerGlubysThrIleGlyAlaThrLeuSerSerLysPhePheGluGlyPhe	122
QY	296	GCATGCTGTGGCTTGGCAATGCTATTGTTTGATGAGCGTATATTATATAGATTGGAGGC	355
Db	123	LysAlaValGlyPheAlaPheMetPheMetAlaIleValValTyLeuTyPheArgAsn	142
QY	356	GTCAATCGCTTCGGAGCTGTTCTTCG-----AATCTTTTGCTTATCTGGGCA	403
Db	143	ProValProSerGlyAlaIleIleLeuSerAlaLeuSerAspIleIleMetAlaLeuGly	162
QY	404	GCTCTACAGTATTGGATGCGCCACCTCACTCTGAGGACTGCTGGGATTGTTCTTGCT	463
Db	163	AlaMetSerLeuLeuGlyIleGluLeuSerSerAlaThrIleAlaAlaLeuMetVal	182
QY	464	ATGGGATGCGCCTAGATGCMAATGTCTT-----GTATTCGAAAGAATCCGA	511
Db	183	IleGlyTySerValAspSerAspIleLeuLeuThrThrArgValLeuLysArgLeuThr	202
QY	512	GAGCAATTTTATGTCTCAAGCTCTTAAATCTGTAGAAAAGGATATACC-----	565
Db	203	LysSerPhe-----AspGluThrValLysGluAlaMetLysThrGlyLeuThrMetThr	220
QY	566	-----AAGCTTTTGGAGCATT	583
Db	221	LeuThrThrIleThrAlaMetLeuIleLeuLeuIleValValLysLeuPheIleProVal	240
QY	584	TTTGAT-----TCTAACTTGACTACAGTATTGGCTCAGCACTTCTT-----	625
Db	241	AlaAspIleLeuAlaAsnIleAlaThrValLeuIleLeuAlaLeuIleAlaAspIleIle	260
QY	626	-----TTCTTCTAGATACAGGCGCTATTAAA	652
Db	261	AsnThrThrTrpLeuAsnAlaGlyIleLeuLys	271
RESULT 37			
YEGN_ECOLI			
ID	YEGN_ECOLI	STANDARD;	PRT; 1040 AA.
AC	P76398; O08005;		
DT	15-JUL-1998 (Rel. 36, Created)		
DT	15-JUL-1998 (Rel. 36, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Hypothetical protein yegN.		
GN	YEGN OR B2075		
OS	Escherichia coli.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Escherichia		
ON	NCBI_TaxID=562;		
OR	[1]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN-K12 / MG1655;		
RX	MEDLINE=97426617; PubMed=9278503;		
RA	Blattner M., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,		
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,		
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,		
RA	Mau B., Shao Y.;		
RT	"The complete genome sequence of Escherichia coli K-12.";		
RL	Science 277:1453-1474 (1997).		
RL	[2]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN=K12;		

RA	MEDLINE=97251358; PubMed=9097040;
RA	Itochi T., Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA	Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA	Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA	Nashimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA	Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA	Yamanoto Y., Horiuchi T.;
RT	"A 460-Rb DNA sequence of the Escherichia coli K-12 genome
RT	corresponding to the 40.1-50.0 min region on the linkage map.";
RL	DNA Res. 3:379-392(1996).
CC	-1- FUNCTION: COULD BE A DRUG EFFLUX PUMP (BY SIMILARITY).
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC	(potential).
CC	-1- SIMILARITY: BELONGS TO THE ACRB/ACRD/ACRF FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; AE000297; AAC75136.1; -;
DR	EMBL; D90845; BAA15929.1; -;
DR	EMBL; D90846; BAA15931.1; -;
DR	ECOGene; EGI4057; YegN.
DR	Inter-Pro; IPRO01036; Acrflvin_res.
DR	Pfam; PF00873; ACR tran; 1.
DR	PRINTS; PR00702; ACRFLVINRP.
KW	Hypothetical protein; Transmembrane; Inner membrane; Transposon;
KW	Complete proteome.
FT	DOMAIN 1 20 CYTOPLASMIC (BY SIMILARITY).
FT	TRANSMEM 21 39 1 (BY SIMILARITY).
FT	DOMAIN 40 342 PERIPLASMIC (BY SIMILARITY).
FT	TRANSMEM 343 362 2 (BY SIMILARITY).
FT	DOMAIN 363 368 CYTOPLASMIC (BY SIMILARITY).
FT	TRANSMEM 369 388 3 (BY SIMILARITY).
FT	DOMAIN 389 394 PERIPLASMIC (BY SIMILARITY).
FT	TRANSMEM 395 416 4 (BY SIMILARITY).
FT	DOMAIN 417 443 CYTOPLASMIC (BY SIMILARITY).
FT	TRANSMEM 444 462 5 (BY SIMILARITY).
FT	DOMAIN 463 475 PERIPLASMIC (BY SIMILARITY).
FT	TRANSMEM 476 498 6 (BY SIMILARITY).
FT	DOMAIN 499 536 CYTOPLASMIC (BY SIMILARITY).
FT	TRANSMEM 537 555 7 (BY SIMILARITY).
FT	DOMAIN 556 865 PERIPLASMIC (BY SIMILARITY).
FT	TRANSMEM 866 885 8 (BY SIMILARITY).
FT	DOMAIN 886 891 CYTOPLASMIC (BY SIMILARITY).
FT	TRANSMEM 892 911 9 (BY SIMILARITY).
FT	DOMAIN 912 917 PERIPLASMIC (BY SIMILARITY).
FT	TRANSMEM 918 939 10 (BY SIMILARITY).
FT	DOMAIN 940 967 CYTOPLASMIC (BY SIMILARITY).
FT	TRANSMEM 968 986 11 (BY SIMILARITY).
FT	DOMAIN 987 999 PERIPLASMIC (BY SIMILARITY).
FT	TRANSMEM 1000 1022 12 (BY SIMILARITY).
FT	DOMAIN 1023 1040 CYTOPLASMIC (BY SIMILARITY).
SO	SEQUENCE 1040 AA; 112077 MW; 195E3D989C7AC6F0 CRC64;

```

Alignment Scores:
Pred. No.: 0.0112 Length: 1040
Score: 129.00 Matches: 102
Percent Similarity: 38.45% Conservative: 81
Best Local Similarity: 21.43% Mismatches: 180
Query Match: 3.89% Indels: 114
DB: 1 Gaps: 18

US-09-868-987-1 (1-1864) x YEGN_ECOLI (1-1040)

Qy 308 TTGGCAATGCTTATTGTTTGATGCGGTATATTAGATT-----GGAGGC 355
      : : : : : : : : : : : : : : : :
Db 349 MetAlaIeAlaLeuValMetIleIleTyLeuPheLeuArgAsnIleProAlaThr 368

```



```

QY 356 CTCATCGCTTGGGAGCTGTTCTTCTGAACTTTTGTATCTGGGAGCTCTACAGAT 415
Db 369 ILeileProGlyValAlaValProLeuSerLeuIleGlyThrPheAlaValMetAlpHe 388
QY 416 TTGATGAGCCACCTACCTCTTGCAGACTCGCTGGAGATGTTGTTGCTATGAGATGACC 475
Db 389 LeuSpHeSerIleAsnLeuThrLeuMetAlaLeuThrIleAlaThrGlyPheVal 408
QY 476 GTAGATGCAATGTTCTTGATTCGAAAGAATC-----CGAGAGGAATTT 520
Db 409 ValAspAspAlaIleValIleGlyAsnIleSerArgTyrIleGlyIleGlyVal 428
QY 521 TTATGTCCTCAAGCTTAAAAATCTGTAGAAAAAGATACCAAGCTTTGGAGCC 580
Db 429 ProLeuAlaAlaIleValGlyAlaGlyIleGlyPheThrIleIle----- 445
QY 581 ATTTTGTATCTTAACCTTACTACAGTATGAGCTTCAGCAGCTCTTTCTCTGATGACA 640
Db 446 -----SerLeuThrPheSerLeuIleAlaValLeuIleProLeuLeuPheMet 461
QY 641 GGGCTATTAAGG-----TTTGTCTTGACATGATTTTGAAGATTTTC 685
Db 462 GlyAspIleValGlyArgLeuPheArgGlyPheAlaIleThrLeuAlaValAlaIleLeu 481
QY 686 TCTTCAATGTTTACGGCTTTTTCATGACTAAATTTTCTTCACGCTCGATGATGATAG 745
Db 482 IleSerAlaValSerLeuThrLeuThrPro-----MetMetCysAlaArgMet 498
QY 746 ACCCAACATCACAGCTTCATATGATGATGAATGATCGTGGAGATTAACATGATTTCTTG 805
Db 499 LeuSerGlnGlySerLeuArgGlyGlnAsnArgPheSerAlaSerGlyIleMetPhe 518
QY 806 -----AGAGATGCAAAAAATCTT-----TGCGCT 829
Db 519 AsparGlyIleAlaAlaIleTyrGlyArgGlyLeuAlaValIleAsnIleProThrLe 538
QY 830 GTTTCGTGAAGTGTGTTCTTTAGGTGCGTGTCTCGGGTGTGAGCCTGGAATTC 889
Db 538 uThrLeuSerVal-----AlaLeuSerThrLeuLeuLeuSerValLe 552
QY 890 GTTTTGGATGATTTTAAAGAGAGGTATGCTTACCT-----TTAATCAAAA 940
Db 552 uLeuThrValPheIleProGlyGlyPhePheProValGlnAspArgIleIleGlnGly 572
QY 941 GAGCATGCGCATCAGGAGATGTTCTCAAAATGCGTGGCAAAATGTCATTAACACAGAA 1000
Db 572 YThrLeuGlnAlaPro----- 577
QY 1001 GCTGCTCTTCTCTAGAGACTTCGCTATTCAAATTTGATCTTCAGAAAGATCAAA 1060
Db 578 -----GlnSerSerPheAlaAsnMetAlaGlnArgGlnArgGlnVala 592
QY 1061 ATCTATTTAGTAAAGCTTTAAGCTATATCTACTAGACATACAGAGCTCTCTCTAAAA 1120
Db 592 AlaAspValIleLeuGlnAspProAlaValGlnSer----- 604
QY 1121 TTAACATCATGAGCTGCGTATTGTGGATGTTGTGCAAAACAGGCTAGATTTCTC 1180
Db 605 -----LeuThrSerPheValGlyValAspGlyTh 614
QY 1181 TAGGAAACTTAACGA-----AAGCAAAATTTTGGTCAAAAGTAAAG 1224
Db 614 rAsnProSerLeuAsnSerAlaArgLeuGlnIleAsnLeuAspProLeuAspGlyArgAs 634
QY 1225 CAGCAAACTATACGAAGAAAG-----CGTTATCAGCGCAGCAATCGGGCTTTAGAGCTTT 1281
Db 634 pAspArgValGlnIleValIleAlaArgLeuGlnThrAlaValAspIleValProGlyVa 654
QY 1282 GGCATCATCTTCTG-----CTCTATGATGATTTGCGCTT 1314
Db 654 lAspLeuPheLeuGlnProThrGlnAspLeuThrIleAspThrGlnValSer---ArgTh 673
QY 1315 TGAATGGCAATATGCTTTCAAGTGGCGTATGCGCTTTAATCATGACCTTTTGGCTACCTG 1374

```

```

Db 673 rGlnTyrGlnPheThrLeuGlnAlaThrSerLeu-----AspAlaLeuSerThr-- 689
QY 1375 TGCAGCTCTTGTATACACATTTCTTTTGAAGAAAATTCAAATATGATTTGCAAGCCAT 1434
Db 690 -----TrpValProGlnLeuMetGlyIleGlnLeuGlnIle 701
QY 1435 TGTGCTTTAATGACTGATTTGGGGTATTCATTAACAAATCTTGATCATTTTGTATCG 1494
Db 701 uProGlnLeuSerAspValSerSerAspTyrGlnAspGlyLeuValAlaTyrValAs 721
QY 1495 TATTCGTGAAGATCGCCAGCAAGCTGTTTACCCCTATGACATGTTTATGATGATGC 1554
Db 721 nValAspArgAspSerAlaSerArgLeuGlyIleSerMet---AlaAspAlaAspAsnAl 740
QY 1555 CTTTCAAAAGACGTTGACGCGCAGCGTATGACAAACGCTACACT 1600
Db 740 aLeuTyrAsnAlaPheGlyGlnArgLeuIleSerThrIleTyrThr 755

RESULT 38
MM3_MYCTU
ID MM3_MYCTU STANDARD; PRT; 944 AA.
AC OS3657;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative membrane protein mmpl3.
GN MMPL3 OR RV0206C OR MT0216 OR MT033.14C OR MTCY0805.01C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetiales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean A., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischnann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Debroy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE MMPL FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL021928; CAI17328.1; -.
CC EMBL; AE006931; AAK4437.1; -.
CC TIGR; MT0216; -.
CC TubercuList; RV0206C; -.

```



DR InterPro; IPR000731; HMGCR/patch\_5TM.  
 DR InterPro; IPR004869; MMPL.  
 DR Pfam; PF03176; MMPL; 1.  
 DR PROSITE; PS50156; SSD; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 14 34 POTENTIAL.  
 FT TRANSMEM 186 206 POTENTIAL.  
 FT TRANSMEM 210 230 POTENTIAL.  
 FT TRANSMEM 236 256 POTENTIAL.  
 FT TRANSMEM 287 307 POTENTIAL.  
 FT TRANSMEM 315 335 POTENTIAL.  
 FT TRANSMEM 397 417 POTENTIAL.  
 FT TRANSMEM 563 583 POTENTIAL.  
 FT TRANSMEM 587 607 POTENTIAL.  
 FT TRANSMEM 617 637 POTENTIAL.  
 FT TRANSMEM 673 693 POTENTIAL.  
 FT TRANSMEM 699 719 POTENTIAL.  
 FT CONFLICT 756 756 K -> N (IN REF. 2).  
 SQ SEQUENCE 944 AA; 100904 MW; 289DF83ACDBE490D CRC64;

## Alignment Scores:

Pred. No.: 0.0133 Length: 944  
 Score: 128.00 Matches: 126  
 Percent Similarity: 36.01% Conservative: 121  
 Best Local Similarity: 18.37% Mismatches: 238  
 Query Match: 3.86% Indels: 201  
 DB: 1 Gaps: 31

US-09-868-987-1 (1-1864) x MML3\_MYCTU (1-944)

Qy 68 GGATGGCGTATGGCTGTAGTATGACGGTATATATGGTCAGCAGC----- 112  
 Db 107 GlyTrrAla-----GlyTrrLeuArgAlaSerGlnAlaThrGlyMet 120  
 Qy 113 -----CCTATTTTAAACGTCCTCATTGAA----- 136  
 Db 121 AlaThrAlaAspLysLeuTyrThrPheValSerIleProLeuLysGlyAspAspAsp 140  
 Qy 137 -----AATCATGCCAGTGTCTCAGGGAATTTTACCACCGTGAAGTGAGCAA 184  
 Db 141 ThrIleLeuAsnAsnTyrLysAlaIleAla-----ProAspLeuGlnArg 155  
 Qy 195 CTC---GCCTCAGATTTAAATCTGGACGATGCTTTTGTTCGCGAGTCTTCAGTAA 241  
 Db 156 LeuAspGlyGlyThrValLysLeuAlaGlyLeuGlnProValAlaGluAlaLeuThr 174  
 Qy 242 GAGACGATCTCTCTGATCTTGGGAAAAACAATGTACACAGGCATTATCTCAGCATGC 301  
 Db 175 GlyThrIleAlaThrAspGlnArgMetGluVal----- 186  
 Qy 302 TGTGGCTTGGCAATGCTTATGTTTGTATGAGCGTATATTATAGATTGGAGCGTCACTC 361  
 Db 187 -----LeuAlaLeuProLeuValAlaValLeuPheValPheGlyGlyValIle 204  
 Qy 362 GCTTCGGAGCTGTT-----CTTCGT 382  
 Db 205 AlaAlaGlyLeuProValMetValGlyGlyLeuCysIleAlaGlyAlaLeuGlyIleMet 224  
 Qy 383 AATCTTTTGTCTATCTGGCAGCTCTACAGTATTTGGATGGCCACTCACCTTGTTCAGA 442  
 Db 225 ArgPheLeuAlaIlePheGlyProValHisTyrPheAlaGlnProValVal----- 241  
 Qy 443 CTCGGTGGATTGTTCTGCTATGGGATGCGCGTAGATGCAAAATGTTCTTATTTCGNA 502  
 Db 242 -----SerLeuIleGlyLeuGlyIleAlaIleAspTyrGlyLeuPheIleValSer 258  
 Qy 503 AGAATCCGAGAGGATTTTATTTGCTCAAGTCTTAAATAATCTGTAGAAAA----- 556  
 Db 259 ArgPheArgGluLeuAlaGluGlyTyrAspThrGluThrAlaValArgArgThrVal 278  
 Qy 557 -----GGATATACCAAGCTTTTGGAGCCATTTTGTATCTTAATCTGACTACAGTA 607  
 Db 279 IleThrAlaGlyArgThrValThrPheSerAlaValLeu-----Ile 292

Qy 608 TTGCCTCAGCA-----CTTCTTTTCTCTAGATACAGGCGCTATTAAAGGTTTCT 661  
 Db 293 ValAlaSerAlaIleGlyLeuLeuPhe-----ProGlnGlyPheLeu 307  
 Qy 662 TTGACATGATTTTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATGACTAAATTT 721  
 Db 308 LysSerLeuThrTyrAlaThrIleAlaSerValMetLeuSerAlaIleLeuSer----- 325  
 Qy 722 TTTCTTCATGCTGTGGATGAATAAGACCAACATACACAGTTGTCATATGATGAATAAGTTC 781  
 Db 326 -----IleThrValLeuProAlaCys 332  
 Qy 782 GTGGGATA-----AAGCATGATTTCTTGAGAGATGCAAAAAACTT----- 823  
 Db 333 LeuGlyIleLeuGlyLysHisValAspAlaLeuGlyValArgThrLeuPheArgValPro 352  
 Qy 824 -----TGGCGTGTCTTGGAGTGTCTTTCTTTAGGTGGTGGTCT----- 865  
 Db 353 PheLeuAlaAsnTrrLysIleSerAlaAla-----TyrLeuAsnTrrLysAlaAspArg 370  
 Qy 866 -----CTCGGTTTGGAGCTCGAATTCGATTCGTTTGGGA 898  
 Db 371 LeuGlnArgThrLysThrArgGluValGluAlaGlyPheTrrGlyLysLeuValAsn 390  
 Qy 899 ATGGATTTTAAA----- 910  
 Db 391 ArgValMetLysArgProValLeuPheAlaAlaProIleValIleIleMetIleLeuLeu 410  
 Qy 911 -----GGAGGTATGCTTTTACCTTTTAATCCAAAGAG 943  
 Db 411 IleIleProValGlyLysLeuSerLeuGlyIleSerGluLysTyrLeuProThr 430  
 Qy 944 CATGCATCAGCATGTGCTCAATCGTGGCAAGTGTGCATAAACTACAG----- 998  
 Db 431 AsnSerValArgGlnAlaGlnGluGlu-PheAspLysLeuPheProGlyTyrArgThrAs 450  
 Qy 999 -AAGCTGTCTTCTTCTAGACTTTCGTATTCAAACATTTG---GATCTTCAG----- 1049  
 Db 450 nProLeuThrLeuValIleGlnThrSerAsnHisGlnProValThrAspAlaGlnIleAl 470  
 Qy 1050 -----AAAGATCAAAATCTATTATGATGATAAAGCTTT----- 1083  
 Db 470 aAspIleArgSerLysAlaMetAlaIle-GlyGlyPheIleGluProAspAsnAspProA 490  
 Qy 1084 -----AAGCTATACTAAGCAGATACGAGCTCTCTCTAAATTAACG 1126  
 Db 490 laAsnMetTrrGlnGluArgAlaTyrAlaValGlyAlaSerLysAspProSerValArgV 510  
 Qy 1127 ATCATGAGCTGGCTGTTATTTGGGATTTGTCAGAAACAGCGCTAGATTTCTCTACGGA 1186  
 Db 510 alLeuGlnAsnGlyLeuIleAsnProAlaAspAlaSerLysLysLeuThrGluLeuArgA 530  
 Qy 1187 AACTCTAAACGAAACGCAAAATTTTGGTCTC-----AAAGGTAAAGCAGCAAACTATCGAAGAAA 1243  
 Db 530 laIleThrProLysGlyIleThrValLeuValGlyGlyThrProAlaLeuGluLeuA 550  
 Qy 1244 ATGGTTATCAGGCGACCATCGGCTTTTAGGAGCTTTGGCAATCATCTTCTCTATGTG 1303  
 Db 550 spSerIleHisGly-LeuPheAlaLysMetProLeuMetValValIleLeuLeuThrThr 569  
 Qy 1304 AGTTTGGCTTTGAATGGCAATATGCTTTCAGTCCGCTATGCGCTTTTAAATTCATGACCTT 1363  
 Db 570 ThrIleValLeuMetPhe-----LeuAlaPheGlySerValValLeuProIleLysAlaThr 588  
 Qy 1364 TTGGCTACCTGTCAGCTCTTG-----TTTATAGCA 1393  
 Db 589 LeuMetSerAlaLeuThrLeuGlySerThrMetGlyIleLeuThrTrrIlePheValAsp 608  
 Qy 1394 CATTTCTTTTGAAGAAATTCAA-----ATAGATTTTCAAGCCCATTTGGT 1438  
 Db 609 GlyHisPheSerLysTrrLysAsnPheThrProThrProLeuThrAlaProValIleGly 628

QY 1439 GCTTTATGACT---GTAITGGSGTATTCATTAACAATCTTGATCATTTTGATCGT 1495  
 Db 629 LeuilellelealeuvalpneglyleuserThraspyrGlualPheleValSerArg 648  
 QY 1496 ATTGCTAAGATCGCAACGACCTGTTTACCCCTTGCATGTTTATGATGATGCC 1555  
 Db 649 MetValGlunlaaArglunArglMetSerThrGlnGlnAlaIleArgIleGly----- 666  
 QY 1556 CTTCAAAAAGCGTTCAGCCGCAAGCTAATGACAACAGCTACACTATCATGTTTGTGA 1615  
 Db 667 -----ThrAlaAlaThrGlyArgIleIleThrAlaAlaAlaLeu 679  
 QY 1616 ATGCTTTTGTATTAGCGCGCTCTGCTGCTTTAAT-----TTT 1654  
 Db 680 lIeValAlaValAlaIaGlyAlaPheValPheSeraspLeuValMetMetLysTyrLeu 699  
 QY 1655 GCATTATATACCATAGACATGATCTCTTA---GGAACCTATATGCTCTTATATATGCA 1711  
 Db 700 AlaPheGlyLeuMetAlaAlaAlaLeuLeuAspAlaThrValAlaArgMetPheLeuVal 719  
 QY 1712 CCACCTCTGTTG 1723  
 Db 720 ProSerValMet 723  
 RESULT 39  
 MMJ3\_MYCLE  
 ID MMJ3\_MYCLE STANDARD; PRT; 955 AA.  
 AC 006081;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative membrane protein mmj3.  
 GN MMJ3 OR ML2620 OR MLC1622.18C.  
 OS Mycobacterium lepreae.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 NC NCB1\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TW;  
 RX MEDLINE=2118732; PubMed=11234002;  
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Fellwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagsels K., Lacroix C., Maclean A., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrett B.G.;  
 RT Massive gene decay in the leprosy bacillus.";  
 RL Nature 409:1007-1011(2001).  
 CC - SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC - SIMILARITY: BELONGS TO THE MMPL FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: Z95398; CAB08813.1; -;  
 CC EMBL: AL583926; CAC32152.1; -;  
 DR Leptoma; ML2620; -;  
 DR InterPro: IPR004869; MMPL.  
 KW Hypochemical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 14 34 POTENTIAL.  
 FT TRANSMEM 191 211 POTENTIAL.  
 FT TRANSMEM 231 251 POTENTIAL.  
 FT TRANSMEM 292 312 POTENTIAL.

FT TRANSMEM 320 340 POTENTIAL.  
 FT TRANSMEM 402 422 POTENTIAL.  
 FT TRANSMEM 568 588 POTENTIAL.  
 FT TRANSMEM 592 612 POTENTIAL.  
 FT TRANSMEM 622 642 POTENTIAL.  
 FT TRANSMEM 679 699 POTENTIAL.  
 FT TRANSMEM 704 724 POTENTIAL.  
 SQ SEQUENCE 955 AA; 102792 MW; 8AF86562EB91D07B CRC64;  
 Alignment Scores:  
 Pred. No.: 0.0212 Length: 955  
 Score: 125.50 Matches: 135  
 Percent Similarity: 35.09% Conservative: 112  
 Best Local Similarity: 19.18 Mismatches: 244  
 Query Match: 3.79% Indels: 213  
 DB: Gaps: 29  
 US-09-868-987-1 (1-1864) x MMJ3\_MYCLE (1-955)  
 QY 50 CAATATTCGCAAAACCGT-----GATGCGTATGCGCTGATGATTCAC 94  
 Db 96 GlnPheGlnArgAsnAsnSerAsnLysValThrGlyTyrAla----- 109  
 QY 95 GGTATATGATGTCAGACGCCCTATTTAAACGTCCTCAATGAAAATCATGCCAGTGTCTCA 154  
 Db 110 GlyTyrLeuArgAlaSeraspThrThrAsnThrValValGlnGlyMetAlaThrProAsp 129  
 QY 155 GGGAAATTTACCCACCGGTGAAGTG----- 178  
 Db 130 LysLysTyrThrPheValSerIleProLeuLysGlyAspAspAspThrIleLeuAsn 149  
 QY 179 -----AGCAACTCGCCCTCAGATTAAATCT-----GCA 208  
 Db 150 AsnTyrLysAlaIleAlaProAspLeuGlnLysLeuAspGlyGlyThrValGlnLeuAla 169  
 QY 209 GCGATGCTTTTGTTCGCCGAGGTCTCAGTGAAGACATCTCTGATCTTGGGAAA 268  
 Db 170 GlyLeuAspProIleAlaAsnAlaLeuThr--SerThrIleAlaThrAspGlnAlaArg 188  
 QY 269 AACAATGTACACAGAGCATATCTCAGACGCTGCGCTGGCAATGCTATGTTGTTTG 328  
 Db 189 MetGluVal-----LeuAlaLeuProLeuValAla 198  
 QY 329 ATGAGCGTATATATATGATTGGAGCGTCATGCTTGGAGACTGTTCTTGAATCTT 388  
 Db 199 ValValLeuPheLeuValPheGlyGlyValIleAla----- 210  
 QY 389 TTGCTATCTGGGACGCTACAGTATTTGGATGGCGACCTGACCTTGTCAGAGTCGCT 448  
 Db 211 -----AlaCysLeuProValMetValGlyGlyLeuSerIleAlaGlyAlaLeu 226  
 QY 449 GGGATT----- 454  
 Db 227 GlyIleAlaArgPheIleAlaLeuPheGlyProValHisPhePheAlaGlnProValAl 246  
 QY 455 ---GTTCTTCATATGGGATGCGCTGATGCAATGTTCTTATTCGAAAGATCCGA 511  
 Db 247 SerLeuIleGlyLeuGlyIleAlaValAspTyrGlyLeuPheValValSerArgPheArg 266  
 QY 512 GAGGAATTTTATATGCTCAAACTTAAATAATCTGTGAGAAA----- 556  
 Db 267 GluGlnIleAlaGluGlyTyrAspThrGlnAlaAlaValArgArgThrValMetThrAla 286  
 QY 557 GGAATATACCAAGCTTTTGGAGCGCATTTTGAATCTTAATCTGACTACAGATGAGCTCA 616  
 Db 287 GlyArgThrValThrPheSerAlaValLeu-----IleAlaAlaSerGly 301  
 QY 617 GCACTTCTTTCTTCTCATATACAGGCGCTATTAAGGCTTGCTTGACATTGAT--- 673  
 Db 302 AlaSerLeuLeuLeuLeuProGlnGlyPheValLysSerLeuThrTyrAlaLeuIleAla 321  
 QY 674 -----TTAGAATTTTC 685

Db 322 AlaValThrLeuAlaAlaLeuSerIleThrLeuLeuProAlaCysLeuAlaIleLeu 341  
Qy 686 TCTTCAATGTTTACGGCTCTT---TTCAAGACTAAATTTTCTTCATGCTGTGGATGAAT 742  
Db 342 AlaLysHisValAspAlaLeuGlyValArgThrLeuPheArgValProLeuArgAsn 361  
Qy 743 AAGACCAACATACACAGCTTCATATGATGAATAGTTTCGTGGGATAAAGCATGATTC 802  
Db 362 TrpArgMetSer-HisAlaCysLeu-----AsnTrp-----LeuAl 373  
Qy 803 TTGAGAGGATCAAAAAACT-----TTGGCTGTTTCTGGAAGT 841  
Db 373 a-AspArgLeuGlnLysThrLysThrArgGluGluValGluAlaGlyPheTrpGlyLysL 393  
Qy 842 GTTTTCTTTAGTTGGCTCTCGGTTTGGAGCTCGAATTCGTTTGGGAATG 901  
Db 393 euValAsnPheValMetLysArgProLeuValPheAlaIleProIle-ValIleGlyMet 412  
Qy 902 GATTTTAAAT-----GGAGGTATGCTTTTACCTTTAAT 934  
Db 413 IleLeuLeuValIleProLeuGlyAsnLeuSerPheGlyGlyMetSerGluLysTrpLeu 432  
Qy 935 CCAAAGAGCATGCGATCAGCGATGTTCTCAAAATGCGTGCAGAAAGTTGTGCATAAACTA 994  
Db 433 ProProAsnAsnAlaValArgGlnSerGlnGluHis-PheAspGlnLeuPheProGlyTy 452  
Qy 995 CAGG-----AAGCTGTCTTCTCTAGAGACTTCGCTATTTCAAAATTTGGATTTCA 1048  
Db 452 rArgThrAsnProLeuThrLeuValIleGlnThr-SerAsnHisGlnProValThrAspGl 472  
Qy 1049 GAAA-----AGATCAAAATCTATTTAGTGTAGTAAGCTTTAAGCTATCTAAG 1096  
Db 472 nGluIleAlaAspIleArgSerLysAlaMetAlaIleSer-GlyPhe-----487  
Qy 1097 CAGATACGAGCTCTCTCTCAAAATTAACGATCATGAGCTGCGTATTGTGG-----1149  
Db 488 -----IleGluPro-----AspAsnAsnTyrrValAsnMetTrpGlnGluA 501  
Qy 1149 -----1149  
Db 501 rgThrValAlaProGlyAlaSerLysAspProSerValArgValLeuGlnAsnGlyLeuI 521  
Qy 1150 -----GATTGTGTGAGAAACAGGCTAGATTTCTCTACGGAACCTCAAACGA 1198  
Db 521 leAsnProAsnAspAlaSerLysLysIleAsn-----GluLeuArgSerIleThrProp 539  
Qy 1199 AACGCAAAATTTTGTGTCAAA---GTTAAGCAGCAAACTATCGAAGAAATCGGTTATCAG 1255  
Db 539 roLysGlyLeuThrValSerValGlyGlyThrProAlaLeuGluGlnAspSerIle-His 558  
Qy 1256 GCGACCATCGGGCTTTTAGGAGCTTTTGGCAATCATCTTGTCTATGTGAGTTTGGCTTT 1315  
Db 559 SerLeuValAlaGlnAlaProLeuMetValIleMetLeuIleThrThrMetLeuLeu 578  
Qy 1316 GAATGGCAATATGCTTTCAGTGGCGTATGCGCTTAATTCATGACCTTTTGGCTACCTGT 1375  
Db 579 MetPhe---LeuAlaPheGlySerPheValLeuProIleLysAlaAlaValMetSerAla 597  
Qy 1376 GCAGTCTGTTTATAGCACATTTCTTTTGAAGAAATTCAAATAGAT-----1423  
Db 598 LeuThrLeuGlySerThrMetGlyIleLeuThrTrpIlePheValAspGlyHisLeuSer 617  
Qy 1424 -----TTGCAAGCCATTTGGTGTCTTTAATGACTGTA 1453  
Db 618 LysTrpLeuAsnPheThrProThrProLeuMetValValIleIleAlaLeuValValAla 637  
Qy 1454 TTGGGTATTCAATAACAATCTTTCATCATTTT-----GATCGATTTCGGAAGAT 1507  
Db 638 ValGlyTyrglyLeuAlaThrAspTyrglyValPheLeuValSerArgMetValGluAla 657  
Qy 1508 CGCAACGAGCACTGTTTACCCTCATGCTTTTGTATGATGATGCTTCAAAAGACG 1567  
Db 658 ArgAlaGluSerMetSerThrGlnGluAlaValArgIleGly-----671

Qy 1568 TTCAGCCGCAACGTAATGACAAACAGCTACAACTCTATCAGTTTGTAAATGCTTTTGT 1627  
Db 672 -----ThrAlaSerThrGlyArgLeuIleThrAlaAlaAlaLeuValLeuAlaVal 688  
Qy 1628 ATAGCGGCTCTCTGCTTTAAT-----TTGCAATTTATTATG 1666  
Db 689 ValAlaGlySerPheValPheSerAspLeuValMetMetLysTyrrLeuAlaPheGlyLeu 708  
Qy 1667 ACCATAGGGATTTCTTCTA---GGAACTTTATCGTCTCTTTATATTGCACCACTCTGTG 1723  
Db 709 MetAlaAlaLeuLeuLeuAspAlaThrValValArgMetPheLeuValProSerValMet 728  
RESULT 40  
YHIV\_ECOLI  
ID YHIV\_ECOLI STANDARD; PRT; 1037 AA.  
AC P37637;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein yhiv.  
GN YHIV OR B3514.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
SEQUENCE FROM N.A.  
STRAIN=K12 / MG1655;  
RX MEDLINE=94316500; PubMed=8041620;  
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
region from 76.0 to 81.5 minutes."  
RL Nucleic Acids Res. 22:2576-2586(1994).  
CC -!- FUNCTION: COULD BE A DRUG EFFLUX PUMP (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
(potential).  
CC -!- SIMILARITY: BELONGS TO THE ACRB/ACRD/ACRF FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announcement/>  
or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
CC EMBL; U00039; AAB18490.1; -;  
DR EMBL; AB000427; AAC76539.1; -;  
DR EcoGene; EGI2241; Yhiv.  
DR InterPro; IPR001036; Acrflvin\_res.  
DR InterPro; IPR004764; HAE1.  
DR Pfam; PF00873; Acr\_tran; 1.  
DR PRINTS; PR00702; ACRFLVINRP.  
DR TIGRFAMs; TIGR00915; 2A0602; 1.  
KW Hypothetical protein; Transmembrane; Inner membrane; Transport;  
KW Complete proteome.  
FT DOMAIN 1 9 CYTOPLASMIC (BY SIMILARITY).  
FT TRANSMEM 10 28 1 (BY SIMILARITY).  
FT DOMAIN 29 339 PERIPLASMIC (BY SIMILARITY).  
FT TRANSMEM 340 359 2 (BY SIMILARITY).  
FT DOMAIN 360 365 CYTOPLASMIC (BY SIMILARITY).  
FT TRANSMEM 366 385 3 (BY SIMILARITY).  
FT DOMAIN 386 391 PERIPLASMIC (BY SIMILARITY).  
FT TRANSMEM 392 413 4 (BY SIMILARITY).  
FT DOMAIN 414 441 CYTOPLASMIC (BY SIMILARITY).  
FT TRANSMEM 442 460 5 (BY SIMILARITY).  
FT DOMAIN 461 473 PERIPLASMIC (BY SIMILARITY).  
FT TRANSMEM 474 496 6 (BY SIMILARITY).  
FT DOMAIN 497 536 CYTOPLASMIC (BY SIMILARITY).  
FT TRANSMEM 537 555 7 (BY SIMILARITY).  
FT DOMAIN 556 580 PERIPLASMIC (BY SIMILARITY).  
FT TRANSMEM 581 871 8 (BY SIMILARITY).



```

Db 875 TyrAlaLeuSerLeuValValPheLeuAlaLeuAlaLeuTyrGluSerTrpSer 894
Oy 1325 TATGCTTTCAGTCCGTA-----TGCCTCTTAATTCATGACCTTTTGGCTACC 1372
Db 895 IleProPheSerValMetLeuValProLeuGlyValValGlyAlaLeuLeuAlaThr 914
Oy 1373 TGTGCGCTCTGTTTATAGCACATCTCTTTTGAAGAAATTCAAATAGATTGCAAGCC 1432
Db 915 -----AspLeuArgGlyLeuSerAsnAspValTyrPhe 925
Oy 1433 ATTGGTGCTTTAATGACTGTATTGGGTATTTCATTAACAATCTTTGATCATTTTGGAT 1492
Db 926 GlnValGlyLeuLeuThrIleGlyLeuSerAlaLysAlaIleLeuValGlu 945
Oy 1493 CGTATTGCTGAAGATGCCAAGCAAGCTGTTTACCCCTATGCTATGTTTAAATGAT 1552
Db 946 PheAlaValGluMetGlnLysGluGlyLysThrProIleGluAlaIleGluAla 965
Oy 1553 GCCCTTCAAAAGAGCTTCAGCGGCAGCGTATGACACAGCTACACTCTATCAGTTTGT 1612
Db 966 AlaArgMetArgLeuArgProIleLeuMetThrSerLeuAlaPheIleLeuGlyValLeu 985
Oy 1613 TTAATGCTTTGTTTATAGGC---GGCTCCTCTGCTCTTAAATTTTGCATTTTATGACC 1669
Db 986 ProLeuValIleSerHisGlyAlaGlySerGlyAlaGlnAsn----- 999
Oy 1670 ATAGGGATTTCTTAGGAACCTTATCGCTCTCTCTATATTTGACCACTCTCTGTTGTTT 1729
Db 1000 --AlaValGlyThrGlyValMetGlyMetPheAlaAlaThrValLeuAlaIleTyr 1018
Oy 1730 ATGCTC 1735
Db 1019 PheVal 1020

RESULT 41
MMLB MYCLE
ID MMLB MYCLE STANDARD; PRT; 1014 AA.
AC 006079;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative membrane protein mmp111.
GN MMLB11 OR M2617 OR MLC622.16C.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridia (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Biglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RL "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE MMLB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; Z95398; CAB08803.1; -.
DR EMBL; AL583926; CAC32149.1; ALT_INIT.
DR Leproma; ML2617; -.
DR InterPro; IPR000731; HMGCR/patch_5TM.
DR InterPro; IPR004869; MMLP.
DR Pfam; PF031176; MMLP; 1.
DR PROSITE; PSS0156; SSD; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 13 33
FT TRANSMEM 188 208 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 235 255 POTENTIAL.
FT TRANSMEM 279 299 POTENTIAL.
FT TRANSMEM 311 331 POTENTIAL.
FT TRANSMEM 373 393 POTENTIAL.
FT TRANSMEM 409 429 POTENTIAL.
FT TRANSMEM 530 550 POTENTIAL.
FT TRANSMEM 560 580 POTENTIAL.
FT TRANSMEM 598 618 POTENTIAL.
FT TRANSMEM 649 669 POTENTIAL.
FT TRANSMEM 671 691 POTENTIAL.
SQ SEQUENCE 1014 AA; 109875 MW; 5BB0D797AFF58D79 CRC64;

Alignment Scores:
Pred. No.: 0.0256 Length: 1014
Score: 124.50 Matches: 103
Percent Similarity: 33.22% Conservative: 85
Best Local Similarity: 18.20% Mismatches: 203
Query Match: 3.76% Indels: 175
DB: 1 Gaps: 24

US-09-868-987-1 (1-1864) x MMLB MYCLE (1-1014)
Oy 374 GTTCTCTGAATCTTTTGTCTTCTGGCGAGCTCTACAGTATTGGTGGCCACTCACC 433
Db 189 IleLeuIleValLeuLeuAlaValPheGlySerLeuAlaAlaAlaValProLeuAla 208
Oy 434 TTGTCAGGACTCGCTGGGATTGTTCTTGTCTATGGG----- 469
Db 209 Leu---GlyValCysThrValValValThrMetGlyLeuValAspLeuValSerMetHis 227
Oy 470 -----ATGCCGCTA 478
Db 228 ThrIleMetSerValPheValThrSerThrValSerMetPheGlyIleAlaLeuAlaVal 247
Oy 479 GATGCAATGTTCTGTGATTCGAAAGATCCGAGAGCAATTTTATTGTCTCAAGCTTT 538
Db 248 AspTyrSerLeuPheIleLeuMetArgPheArgGluLeuArgSerGlyArgGlnPro 267
Oy 539 AAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTATTCTAACTTG 598
Db 268 GlnGluAlaValAspAlaAlaMetAlaThrSerGlyLeuAlaValLeuSerGlyMet 287
Oy 599 ACTACAGTATTGGCTCAGCAGCTCTTTTCTCTAGATACAGGGCCCTATTAAAGGTTT 658
Db 288 ThrValIleAlaSerLeuThrGlyIleTyrLeuIleAsnThrAlaAlaLeuLysSerMet 307
Oy 659 GCTTTGACATGATTTTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATGACT--- 715
Db 308 AlaThrGlyAlaIleLeuAlaValAlaIleAlaMetLeuAlaSerIleThrLeuThrPro 327
Oy 716 -----AAATTTTCTTCATGCTCTGGATG 739
Db 328 AlaAlaLeuAlaThrPheGlyArgAlaAlaValLysArgSerValLeuMetHisTrpSer 347
Oy 740 AATAAGACCAACATACACAGATTCATATGATGAATTAAGTTCTGGGGG---ATAAGCAT 796
Db 348 GlnArgSerGluCysThrGlnSerLeuPheThrThrArgTrpValGlyTrpValMetHis 367
Oy 797 GATTTCTTGAGAGATGCCAAAAAATCTTGG-----GCTGTTTCTGGAAGTGT 844
Db 368 -----ArgProTrpIleSerAlaSerAlaAlaSerThrIleLeu 380

```

QY 845 TTTCTTTAGGTGGCTTCTCTGGGTTTGAGCGCTTGAATCCGTTTGGGAATGAT 904  
 Db 381 TlellelelathPrvValThrSerMetLeuGlyAsnSerLeuValGln-- 399  
 QY 905 TTTAAGAGAGGTATGCTTTTCTTATTCCTTAAAGAGCATGCGATCAGCAATGTTGCT 964  
 Db 400 -----PheAsnSerSerHis----- 404  
 QY 965 CAAATGCGTGGCAAGTTGTGCATTAACAGAGAGCTGTTCTTTCTTACAGACTTC 1024  
 Db 405 GlutlelrgalaglyValalaalalaalaglnaleuGlyProglYalaleu----- 422  
 QY 1025 CGATTTCAAACATTGGATCTCTTGAAGAAAGATCAAAATCTTTTATGATGAAGCTTTA 1084  
 Db 423 -----GlyProvalglInvalleuilethPrProAspPro-- 435  
 QY 1085 AGTATTAAGTAAAGATACGAGCTCTCTCTTAAATTAAGATCATGAGCTGCGTTAT 1144  
 Db 436 -----AsnthrglnAlaSerSerProlyshlsarglnthrllegYala----- 450  
 QY 1145 TGTGGGATTGTTGTCAGAACAG- 1168  
 Db 451 -----IleArgAsnArgMetLeuGlnAlaLysAsnValMetSerValAlaPro 466  
 QY 1169 CCTAGATTCTCTACGGAAC----- 1189  
 Db 467 ProglInPheAlaAspAsnAsnCySerAlaLeuLeuGlyAlaValLeuSerValAspPro 486  
 QY 1190 -----CTAAACGAAACGCAAAATTTGTCTCAAGTAAAGTAAAGCAACTATCGAAG 1240  
 Db 487 GlusAspLeuGlyAlaArgGlnThValAspIthPrMetArgthrglu-----LeuPro 503  
 QY 1241 AAAATGCGTTATCGAGCCAGCCATCGGCTTTTGAAGCTTTGGCATGATC----- 1291  
 Db 504 LysValProglYalalaAlaAlaValAlaValAlaGlyProthrlAlaLeuIleAsnAspPhe 523  
 QY 1292 -----TTCCTCAT 1300  
 Db 524 AspAspArgValAlaLysThGlnProLeuMetLeuValPheValAlaLeuIleAlaPro 543  
 QY 1301 GTGAGTTTGGCTTTGATGCAATATGCTTTGATGCGCGTATGCTTATTCATGAC 1360  
 Db 544 ValMetLeuLeuIleSerIleArgSerValPheLeuAlaPheLysGlyValLeuMetThr 563  
 QY 1361 CTTTGGCTACTGTGCAGTC----- 1381  
 Db 564 LeuLeuSerValAlaAlaAlaLysGlySerLeuValMetValPheGlnTrpGlyTrpLeu 583  
 QY 1382 -----TTGTTATAGCAGCATTTCTTTTGAAGAAATTCAAATAGATTTCAGAGCAT 1435  
 Db 584 GlusAsnLeuGlyPheThHis-----IleAsnSerIleAspSerThValProProLeu 601  
 QY 1436 GGTCTTTTATGACTGTATGGGATTCATTAACAATACTTTGATCATTTTGAATCGT 1495  
 Db 602 ValLeuAlaMetThr-----PheGlyLeuSerMetAspArgGluIlePheLeuLeuThrArg 620  
 QY 1496 ATTGCTGAAGATCGCCAGCAACCTGTTTACCCCTATGATGTTTATGTTAATGATGCC 1555  
 Db 621 IleArgGlu-----ArgPheLeuGlnThrgHis-----ThrArgAspAla 634  
 QY 1556 CTTCAAAAAGAGCTTACGCGACGATATGACAAAGTATACATCTATCATGTTTGTATA 1615  
 Db 635 ValAlaTrgIyValSer--ThrSerAlaArgThrllethSerAlaAlaLeuIleLeu 653  
 QY 1616 ATGCTTTTGTATAGGCGGCTCTGTCTTATATTGCA-----TTTATTATG 1666  
 Db 654 IleAlaValPheValGly-----PheAlaPheAlaGlyMetProLeuValAla 669  
 QY 1667 ACCATAGGATCTTCTAGGA-----ACTTATCGTCTCTTAT 1705  
 Db 670 GlutleGlyValAlaLysAlaValAlaIleAlaValAspValThrlAlaValArgLeuVal 689

QY 1706 ATTCACCACTCTGTG 1723  
 Db 690 LeuValProThrLeuMet 695  
 RESULT 42  
 ID YEGO ECOLI STANDARD; PRT; 1025 AA.  
 AC P6359; 008006;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein yego.  
 GN YEGO OR B2076.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxId=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SRRAIN=K12 / MG1655;  
 RA Blattner F.R., Plunkett G. III, Mayhew G.F., Perna N.T.,  
 RA Glasner F.D.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SRRAIN=K12;  
 RA MEDLINE=97251358; PubMed=9097040;  
 RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,  
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima H.,  
 RA Saito N., Sempel G., Seki Y., Sivaraman S., Tagami H.,  
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horuchi T.;  
 RA "A 460-Pb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 40,1-50.0 min region on the linkage map."  
 RL DNA Res. 3:379-392(1996).  
 CC -1- FUNCTION: COULD BE A DRUG EFFLUX PUMP (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE ACRB/ACRP FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AE000297; AAC75137.1; -.  
 DR EMBL: D90845; BAA15930.1; -.  
 DR EMBL: D90846; BAA15932.1; -.  
 DR Ecogene; EGI4058; Yego.  
 DR InterPro; IPR001036; Acrflvin\_res.  
 DR Pfam; PF00873; ACR\_tran; 1.  
 DR PRINTS; PR00702; ACRIFLAVINRP.  
 KW Hypothetical protein; Transmembrane; Inner membrane; Transport;  
 KW Complete proteome.  
 FT DOMAIN 1 11  
 FT TRANSMEM 12 30  
 FT DOMAIN 31 333  
 FT TRANSMEM 334 353  
 FT TRANSMEM 354 359  
 FT TRANSMEM 360 379  
 FT TRANSMEM 380 385  
 FT TRANSMEM 386 407  
 FT TRANSMEM 408 434  
 FT TRANSMEM 435 453  
 FT TRANSMEM 454 466  
 FT TRANSMEM 467 489  
 FT TRANSMEM 490 526  
 FT TRANSMEM 527 545  
 FT TRANSMEM 546 851  
 FT DOMAIN 546 851  
 CYTOPLASMIC (BY SIMILARITY).  
 1 (BY SIMILARITY).  
 PERIPLASMIC (BY SIMILARITY).  
 2 (BY SIMILARITY).  
 CYTOPLASMIC (BY SIMILARITY).  
 3 (BY SIMILARITY).  
 PERIPLASMIC (BY SIMILARITY).  
 4 (BY SIMILARITY).  
 CYTOPLASMIC (BY SIMILARITY).  
 5 (BY SIMILARITY).  
 PERIPLASMIC (BY SIMILARITY).  
 6 (BY SIMILARITY).  
 CYTOPLASMIC (BY SIMILARITY).  
 7 (BY SIMILARITY).  
 PERIPLASMIC (BY SIMILARITY).









```

Db 376 LeuLeuGlyThrPheGlyValLeuAlaPheGlyPheSerIleAsnThrLeuThrMet 395
Qy 437 TCAGAGCTCGTGGGATGTTCTGCTATGGGARGCGGTAGATGCAAAATGTTCTGTA 496
Db 396 PheGlyMet-----ValLeuAlaIleGlyLeuValAspAlaIleValVal 412
Qy 497 TTCGAAGAATCCGAGAGGAATTTTATTGCTCAAGTCTT-----AAAAATCTGTA 550
Db 413 ValGluAsnVal---GluArgValMetAlaGluGlyLeuSerProArgIleAlaAla 431
Qy 551 GAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGTGATCTTAACCTGACTACAGTATTG 610
Db 432 ArgIysSerMetGlyGlnIleGlnGlyAlaLeuVal-----GlyIleAlaMetValLeu 449
Qy 611 GCCTCAGCATTCTT-----TTCTTCCTAGATACAGGCGCTATT---AAAGGG 655
Db 450 SerAlaValPheLeuProMetAlaPhePheGlyGlySerThrGlyValIleIleArgGln 469
Qy 656 TTGCTTTGACATTGATTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATGACT 715
Db 470 PheSerIleThrIleValSerAlaMetAlaLeuSerValIleValAlaLeuIleLeuThr 489
Qy 716 -----AAA 718
Db 490 ProAlaLeuCysAlaThrMetLeuLysProIleGluLysGlyAspHisGlyGluHisLys 509
Qy 719 TTTTCTTCATGCTGTGGATGAATAAGAACCAACATACACAGTTGCATPATGATGAATAAG 778
Db 510 GlyGlyPhePheGlyTrpPheAsnArg-----Met 519
Qy 779 TTCGTGGGATAAGCAGTATTCTTGAGAGATGCAAAAACTTTGG-----826
Db 520 PheLeuSerThrThrHisGlyTyrrGluArgGlyValAlaSerIleLeuLysHisArgAla 539
Qy 827 -----GCTGTTCTCGAAGTGTTTCTTTTAGTTGGCTT 862
Db 540 ProTyrrLeuLeuIleTyrrValIleValAlaGlyMetIleTrpMetPheThrArgIle 559
Qy 863 GCTCTCGGGTT-----874
Db 560 ProThrAlaPheLeuProAspGluAspGlnGlyValLeuPheAlaGlnValGlnThrPro 579
Qy 874 -----874
Db 580 ProGlySerSerAlaGluArgThrGlnValValAspSerMetArgGluTyrrLeuLeu 599
Qy 874 -----874
Db 600 GluLysGluSerSerSerValSerSerValPheThrValThrGlyPheAsnPheAlaGly 619
Qy 875 -----GGA 877
Db 620 ArgGlyGlnSerSerGlyMetAlaPheIleMetLeuLysProTrpGluGluArgProGly 639
Qy 878 GCCTGGAATCCGTTTGGGAATG-----GATTTHAA 910
Db 640 GlyGluAsnSerValPheGluLeuAlaLysArgAlaGlnMetHisPhePheSerPheLys 659
Qy 911 GGAGGGTATGCTTACCTTTAAATCCA-----937
Db 660 AspAlaMetValPheAlaPheAlaProProSerValLeuGluLeuGlyAsnAlaThrGly 679
Qy 937 -----937
Db 680 PheAspLeuPheLeuGlnAspGlnAlaGlyValGlyHisGluValLeuGluAlaArg 699
Qy 938 -----AAAGCATGGC 949
Db 700 AsnLysPheLeuMetLeuAlaAlaGlnAsnProAlaLeuGlnArgValArgProAsnGly 719
Qy 950 ATCAGCGATGTTGCTCAATCGGTGGCAAGT---GTGCATAAACTACAGGAAGCTGGT 1006
Db 720 MetSerAspGluProGlnTyrrLysLeuGluIleAspAspGluLysAlaSerAlaLeuGly 739

```

```

Qy 1007 CTTTCTTCTAGAGACTTCCGT-----ATTCAA 1033
Db 740 ValSerLeuAlaAspIleAsnSerThrValSerIleAlaTrpGlySerSerTyrrValAsn 759
Qy 1034 ACATTGGATCTTCCAGAAAGATCAAA---ATCTATTATT-----1069
Db 760 AspPheIleAspArgGlyArgValIleArgValTyrrLeuGlnGlyArgProAspAlaArg 779
Qy 1070 -----AGTGATAAGCTTTAAGCTTACTACTAGACAGATACGAGCTCTCTCTCTAAAA 1120
Db 780 MetAsnProAspAspLeuSerLysTrpTyrrValArgAsnAspLysGlyGluMetValPro 799
Qy 1121 TTAACGATCATGAGC-----TGG-----CGTTATTGT 1147
Db 800 PheAsnAlaPheAlaThrGlyLysTrpGluTyrrGlySerProLysLeuGluArgTyrrAsn 819
Qy 1148 GGG-----ATTGTTGTCAGAAACAGGCTAGATTCTCTACGGAACTCT 1192
Db 820 GlyValProAlaMetGluIleLeuGlyGluProAlaProGlyLeuSerSerGlyAspAla 839
Qy 1193 AAACGAAACGCAAAATTTTGTCAAGTAAGCAGCAAACTATCGAAGAAATGCGTTAT 1252
Db 840 MetAlaIleValGlu-----GluIleValLysGlnLeuProLysGlyValGlyTyrr 856
Qy 1253 CAGCGGACCATCGGCTT-----1270
Db 857 SerTrpThr---GlyLeuSerTyrrGluGluArgLeuSerGlySerGlnAlaProAlaLeu 875
Qy 1271 -----TTAGAGCTTTGGCAATCATCTTCTCTATGTGAGTTTGGCTTTGATGCAAA 1324
Db 876 TyrrAlaLeuSerLeuLeuValValPheLeuCysLeuAlaAlaLeuTyrrGluSerTrpSer 895
Qy 1325 TATGCTTTCACTGTCGCTA-----TGCCTTTTAATTATGACCTTTTCGCTACC 1372
Db 896 IleProPheSerValMetLeuValValProLeuGlyValIleGlyAlaLeuLeuAlaThr 915
Qy 1373 TGTGCACTCTGTTTATAGCACATTTCTTTTGAAGAAAAATTCAAATAGATTTCGAAGCC 1432
Db 916 Ser-----MetArgGlyLeuSerAsnAspValPhePhe 926
Qy 1433 ATTGCTGCTTTAATGACTGTATGGGCTATTCATTAACAATACTTTTCATCTTTTGCAT 1492
Db 927 GlnValGlyLeuLeuThrThrIleGlyLeuSerAlaLysAsnAlaIleLeuIleValGlu 946
Qy 1493 CGTATTGCTGAAGATCGCAAGCAACCTGTTTACCCTATGCATGTTTATAGTTAAATGAT 1552
Db 947 PheAlaLysGlu-----LeuHisGluGlnGlyLysGlyIleValGluAla 961
Qy 1553 GCCCTTCAAAAGACGTTTCAGCCGC-----ACGGTAATGACAACA---GCTACAACT 1600
Db 962 AlaIleGluAlaCysArgMetArgLeuArgProIleValMetThrSerLeuAlaPheIle 981
Qy 1601 CTATCAGTTTGTTAATGCTTTTGTATAGCGGCTCTCTCTCTCTTTAATTTTGCATTT 1660
Db 982 LeuGlyValValProLeuAlaIleSerThrGlyAlaGlySerGlySerGlnHisAla---1000
Qy 1661 ATTATGACCATAGGATTTCTTAGGAACCTTTATCGTCTCTTTTATATTGCACCACTCTG 1720
Db 1001 -----IleGlyThrGlyValIleGlyMetValThrAlaThrValLeu 1015
Qy 1721 TTGTTGTTTATGGTC 1735
Db 1016 AlaIlePheTrpVal 1020

```

```

RESULT 44
MMLE MYCTU          STANDARD;      PRT;    966 AA.
ID   MMLE MYCTU
AC   OS3653;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative membrane protein mmpL11.

```

GN MMBP1 OR RV0202C OR MT0212 OR MT033.10C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV; PubMed=9634230;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gae S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulten J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Knouri H., Gill J., Mikula A.,  
 RA Bishal W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1 SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1 SIMILARITY: BELONGS TO THE MMBP FAMILY.  
 CC ;  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AL021928; CAA17334.1; -  
 DR EMBL: AE006930; AAK44433.1; -  
 DR TIGR: MT0212; -  
 DR TuberculList; RV0202c; -  
 DR InterPro; IPR000731; HMGCR/patch\_5TM.  
 DR PROSITE; PS50156; SSD; 1.  
 KM Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 13 33 POTENTIAL.  
 FT TRANSMEM 188 208 POTENTIAL.  
 FT TRANSMEM 214 234 POTENTIAL.  
 FT TRANSMEM 235 255 POTENTIAL.  
 FT TRANSMEM 279 299 POTENTIAL.  
 FT TRANSMEM 311 331 POTENTIAL.  
 FT TRANSMEM 373 393 POTENTIAL.  
 FT TRANSMEM 409 429 POTENTIAL.  
 FT TRANSMEM 527 547 POTENTIAL.  
 FT TRANSMEM 557 577 POTENTIAL.  
 FT TRANSMEM 595 615 POTENTIAL.  
 FT TRANSMEM 646 666 POTENTIAL.  
 FT TRANSMEM 668 688 POTENTIAL.  
 SQ SEQUENCE 966 AA; 103502 MW; 3E8BF0327CBEA2DA CRC64;

## Alignment Scores:

Pred. No.: 0.0445 Length: 966  
 Score: 121.50 Matches: 110  
 Percent Similarity: 33.7% Conservative: 92  
 Best Local Similarity: 18.06% Mismatch: 210  
 Query Match: 3.67% Indels: 197  
 DB: 1 Gaps: 25

US-09-868-987-1 (1-1864) x MMBP\_MYCTU (1-966)

QY 281 CAAGCATTTCTCAGATCTCTGGCTTG----- 310  
 Db 163 GlnGlyValLeuSerAlaAlaAlaAlaAsnThrLysHisAspIleAlaAsnAlaGlu 182  
 QY 311 -----GCAATGCTTATGTTTGTATGAGCGGTATATATATAGATTGGAGCG 355  
 Db 183 ArgTyrAsnLeuProIleIleLeuMetValLeuValAlaAlaPhe-----Gly 198  
 QY 356 GTCATGCGCTTCGGAGCGTGTCTTCGAACTT-----TTGCTTATCTGGGCA 403  
 Db 199 SerLeuAlaAlaAlaAlaIleProLeuAlaLeuAlaValCysThrValIleThrMet 218  
 QY 404 GCTCTACAGAT--TTGATGCGCCACTACCTCTGACAGACTCGGTGGATTGTTCTT 460  
 Db 219 GlyLeuValAlaPheValLeuSerMetHisThrThrMetSerValPheValThrSerThrVal 238  
 QY 461 GCTATG-----GGATGCGCGTAGATGCAAAAGTTCTGTATTCGAAAGATCCGA 511  
 Db 239 SerMetHeGlyIleAlaAlaLeuAlaValAspTyrSerLeuHeIleLeuMetArgTyrArg 258  
 QY 512 GAGCAATTTTATTTGCTCAAGCTTTAAAAAATCTGTAGAAAAGATATACAGCGCT 571  
 Db 259 GluGluLeuArgCysGlyArgArgProArgAlaValAlaAspAlaIleMetAlaThrSer 278  
 QY 572 TTGGAGCCATTTTGTATTTCTTAACTTGACTGACATGATATGGCCCTGACACTCTTCTTC 631  
 Db 279 GlyLeuAlaValAlaValLeuSerGlyMetThrValIleLeuSerLeuThrGlyIleTyrLeu 298  
 QY 632 CTGATATACAGGCGCTTAAAGGTTTGTGACATGATTTTGAATTTTCTCTCA 691  
 Db 299 IleAsnThrProAlaLeuArgSerMetAlaThrGlyAlaIleLeuAlaValAlaAla 318  
 QY 692 ATGTTAGCGCTCTTTCATGACT----- 715  
 Db 319 MetLeuThrSerAlaThrLeuThrProAlaValLeuAlaThrPheAlaArgAlaAlaAla 338  
 QY 716 ---AAATTTTCTTCATGCTGCTGATGAAATTAAGCCCAATACACAGTTGCATATGATG 772  
 Db 339 LysArgSerAlaLeuValHisThrSerArgArgProAlaSerThrGlnSerThrPheTyr 358  
 QY 773 AATAAGTCTGGGGATTAAGCATGATTTCTGAGAGATGCAAAAACCTTGG----- 826  
 Db 359 SerArgTyrValGlyTyrPval-----MetArgArgProThrIleThr 372  
 QY 827 GCTGTTTCTGGAAGCTTTTCTTTAGTTGGTTGCT-----CTC 868  
 Db 373 AlaLeuAlaAlaSerThrValLeuValMetAlaAlaProAlaThrLeuMetValLeu 392  
 QY 869 GGGTTTGAGACCTGGAATTCCTGTTTGGGAATGATTTTAAAGAGGATAGCTTTACC 928  
 Db 393 Gly-----AsnSerLeuLeu--ArgGlnPheSerSerHisGluIle-- 406  
 QY 929 TTTAATCAAAAGACATGCGATCAGATGTTGCTCAATGCGTGCG- 976  
 Db 407 -----ArgThrGlyAlaAlaAlaAlaAlaAlaLeuGlyProGlyAlaLeu 422  
 QY 977 -----AAAGTTGCAATAAATAACAGAAAGCTGTTCTTCTTACAGACTTCGCT 1027  
 Db 423 GlyProValGlnValLeuValArgPheAspAlaGlyGlyAlaSerAlaProGlu--His 441  
 QY 1028 ATTCAAAATTTGGATCTTCAAGAAAGATCAAAATCTATTTTGTGATTAAGCTTAAAGC 1087  
 Db 442 SerGlnThrIleAlaAla----- 447  
 QY 1088 TATACTAAGACAGATACAGCGCTCTCTCTAAATTAACATGACATGAGCTGCGTATTTGT 1147  
 Db 448 -----IleArgHisArgIleAlaGlnAlaPro----- 456  
 QY 1148 GGGATTTGTTGTCAAAAACAGCGCTTATTTCTTACGGAACCTTAA----- 1195  
 Db 457 AsnValValSerValAlaProProAlaArgPheAlaAspAspAsnGlySerAlaLeuLeuSer 476



```

FT DOMAIN 360 365 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 366 385 3 (BY SIMILARITY).
FT DOMAIN 386 391 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 392 413 4 (BY SIMILARITY).
FT DOMAIN 414 441 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 442 460 5 (BY SIMILARITY).
FT DOMAIN 461 473 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 474 496 6 (BY SIMILARITY).
FT DOMAIN 497 537 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 538 556 7 (BY SIMILARITY).
FT DOMAIN 557 570 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 571 590 8 (BY SIMILARITY).
FT DOMAIN 591 603 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 604 1026 12 (BY SIMILARITY).
FT DOMAIN 1027 1037 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 1038 1037 A -> G (IN REF. 1).
FT CONFLICT 303 303 V -> E (IN REF. 1).
FT CONFLICT 372 372 A -> D (IN REF. 2).
FT CONFLICT 385 385 G -> P (IN REF. 1).
FT CONFLICT 461 461 S -> PD (IN REF. 1).
FT CONFLICT 665 665 R -> A (IN REF. 1).
FT CONFLICT 763 763 A -> G (IN REF. 1).
FT CONFLICT 778 778 R -> P (IN REF. 1).
SQ SEQUENCE 1037 AA, 113047 MW, 961611ED124FD4E5 CRC64;

```

```

Alignment Scores:
Pred. No.: 0.0783 Length: 1037
Score: 118.50 Matches: 146
Percent Similarity: 30.26% Conservative: 97
Best local Similarity: 18.18% Mismatches: 267
Query Match: 3.58% Indels: 293
DB: 1 Gaps: 33

```

US-09-868-987-1 (1-1864) x ACRD\_ECOLI (1-1037)

```

QY 44 AATGGACAAATTTCTGGAACCGGTGATGGCTGATGAGTATGACGGTATATG 103
   |||||:::|||||:::|||||:::|||||:::
DB 282 AsnGlySerProAlaSerGlyLeuGlyValLeuAlaSerGlyAlaAsnGlyMetAla 301
QY 104 GTCAGCAGCCCTATTAAACGTCCTGAAATCATGCGACGTCTCGAGGAAATT 163
   :::::|||||:::|||||:::|||||:::
DB 302 ThrLeuLeuValLeuAsn-----ArgLeuAspGlyLeuAlaGlnTyrPhe 317
QY 164 ACCCAGCGTGAAGTGAACAACCTGCTCAGATTAAATCTGAGCGATGCTTTGTT 223
   |||||:::|||||:::|||||:::|||||:::
DB 318 ProHieGlyLeuGlyTyrValLeuAlaTyrGlnThrSerPheValLysAlaSerIle 337
QY 224 CCGAGTCTTCAGTGAAGAGACGATCTCTGATCTTGGAAAAAACAATGATACAA 283
   :::::|||||:::|||||:::|||||:::
DB 338 GluAspValVal-----LysThrLeuLeuGln 346
QY 284 GGCATTATCTCAGCATGCTGTGGCTTGCAATGCTTATTTGTTGATGACGATATATTAT 343
   |||||:::|||||:::|||||:::|||||:::
DB 347 AlaIle-----AlaLeuValPheLeuValMetTyrLeuPheLeuGlnAsnPheArgAla 364
QY 344 AGATTGGAGGCGTCAATGCTTGGAGAGCTTCTTCTGAAATCTTTGCTTATCTGGCA 403
   |||||:::|||||:::|||||:::|||||:::
DB 365 ThrLeuLeuProThrIleAlaValProValLeuMetGlyThrPheSerValLeuTyr 384
QY 404 GCTTACAGTATTTGATGATGCCCATCAGCTTTCAGACATCGCTGGGATGTTGCTGCT 463
   |||||:::|||||:::|||||:::|||||:::
DB 385 AlaHieGlyTyrSerValAsnThrLeuThrMetPheAlaMet-----ValLeuAla 401
QY 464 ATGGGATGGCGGTAGATGCAAAATGTTCTTATTCGAAAGATCCGAGAGAAATTTTAA 523
   |||||:::|||||:::|||||:::|||||:::
DB 402 IleGlyLeuLeuValAlaSerPheAlaIleValValGluAsnVal---GluArgIleMet 420

```

```

QY 524 TTGTCGAAGTCT-----AAAAATCGTGAAGAAAGATATACCAAGCTTTTGA 577
   ::|||:::|||||:::|||||:::
DB 421 SerGluGlyLeuThrProArgGluAlaThrArgLysSerMetGlyGlnIleGlnGly 440
QY 578 GCCATT-----TTGATCTCACTTGCATCAGTACGATTTGGCTTCACACTCTTTCTTC 621
   |||||:::|||||:::|||||:::
DB 441 AlaLeuValGlyIleAlaMetValLeuSerAlaValPheValProMetAlaPhePheGly 460
QY 632 CTGATACAGAGCGCTTAT--AAAGGTTTGCTTACATGATTTTGAATTTTCTCT 688
   |||||:::|||||:::|||||:::
DB 461 GlyThrThrGlyAlaIleTyrArgGlnPheSerIleThrIleValAlaAlaMetValLeu 480
QY 689 TCATGTTTACGGCTCTTTCATGACTCAATTTTC----- 724
   |||||:::|||||:::|||||:::
DB 481 SerValLeuValAlaMetIleLeuThrProAlaLeuCysAlaThrLeuLeuLysProLeu 500
QY 725 -----TTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 754
   |||||:::|||||:::|||||:::
DB 501 LysLysGlyGlyIleHisGlyGlnGlyGlyPhePheAlaThrPheAsnGlnMetPheAsn 520
QY 755 ACAAGTTCGATATGATGATTAAGTTCGAGG--ATTAAGATGATTTCTTGAGAGCA 811
   |||||:::|||||:::|||||:::
DB 521 ArgAsnAlaGluArgTyrGlyValAlaLysIleLeuHisArgSerLeuArg--- 539
QY 812 TGCAAAAACTTTGGCTGTT-----TCGGAAGTGTCTTCTT--- 850
   |||||:::|||||:::|||||:::
DB 540 -----ThrIleValIleTyrValLeuLeuLeuGlyGlyMetValPheLeuPhe 555
QY 850 ----- 850
DB 556 LeuArgLeuProThrSerPheLeuProLeuGluAspArgGlyMetPheThrThrSerVal 575
QY 850 ----- 850
DB 576 GlnLeuProSerGlySerThrGlnGlnThrLeuLysValGluGlnIleGlyLys 595
QY 851 -----TTAGGTGGCTGTGCTCGGG 871
   |||||:::|||||:::|||||:::
DB 596 TyrTyrPheThrHisGlyLysAspAsnIleMetSerValPheAlaThrValGlySerGly 615
QY 872 TTGAGCC-----TGAAATTC 889
   |||||:::|||||:::|||||:::
DB 616 ProGlyLysnGlyGlnAsnValAlaArgMetPheIleArgLeuLysAspTyrPseGln 635
QY 890 GTTTGGAAATGATTTTAAAGAGGCTATGCTTTACCTT----- 931
   |||||:::|||||:::|||||:::
DB 636 -----ArgAspSerLysThrGlyThrSerPheAlaIleIleGlyArgAlaThrLys 652
QY 931 ----- 931
DB 653 AlaPheAsnGlnIleLysGluAlaArgValIleAlaSerSerProProAlaIleSerGly 672
QY 931 ----- 931
DB 673 LeuGlySerSerAlaGlyPheAspMetGluLeuGlnAspHisAlaGlyAlaGlyHisAsp 692
QY 932 -----AATCCA----- 937
   |||||:::|||||:::|||||:::
DB 693 AlaLeuMetAlaAlaArgAsnGlnLeuLeuAlaLeuAlaIleGluAsnProGluLeuThr 712
QY 938 -----AAAGAGATGCGATCAGCGATGTTGCTCAAAATGGTGAGCAAGT---GTGCAT 968
   :::::|||||:::|||||:::|||||:::
DB 713 ArgValArgHisAsnGlyLeuAspAspSerProGlnLeuGlnIleAspIleAspGlnArg 732
QY 989 AAACTAAGAGAACCTGCTTCTTCTAGAGACTTCCTG-----ATTCAAACA---TTT 1039
   |||||:::|||||:::|||||:::|||||:::
DB 733 LysAlaGlnAlaLeuGlyValAlaIleAspAspIleAsnAspThrLeuGlnThrAlaTyr 752
QY 1040 GGATCTTCA-----GAAAAGTCAAAATCTATTTTACTGAT 1075
   |||||:::|||||:::|||||:::
DB 753 GlySerSerTyrValAsnAspPheMetAspArgGlyArgValLysValTyrValGln 772
QY 1076 AAAGCTTAAAGCTATCTAAGCAGATACGAGCTCTCTCTAAATTAACGATCAGAGC 1135

```

```
Db 773 AlaAlaProTyr-----ArgMetLeuProAspAspIleAsnLeu 786
||| ||| :||| |||
Qy 1136 TGGCGTTATTGGGATTGTTGTGAGAAACAGCGCTAGATTCTCTACGGAAACTCTAAA 1195
||| ||| ||| ||| :||| |||
Db 787 Trp-----TyrValArgAsnLysAspGlyGlyMetValProPheSerAla 801
||| ||| ||| ||| :||| |||
Qy 1196 CGAAACGCAAAATTTTGGTCAAGGTAAAGCAGCAAACTATCGAAGAAATCGGTTATCAG 1255
||| ||| ||| ||| :||| |||
Db 802 PheAlaThrSerArgTrpGluThrGlySerProArgLeuGluArgTyrAsnGlyTyrSer 821
||| ||| ||| ||| :||| |||
Qy 1256 CCGACC----- 1261
||| |||
Db 822 AlaValGluIleValGlyGluAlaAlaProGlyValSerThrGlyThrAlaMetAspIle 841
||| ||| ||| |||
Qy 1262 -----ATCGGGCTT----- 1270
||| |||
Db 842 MetGluSerLeuValLysGlnLeuProAsnGlyPheGlyLeuGluTrpThrAlaMetSer 861
||| ||| ||| |||
Qy 1271 -----TTAGGAGCTTTGGCA 1285
||| |||
Db 862 TyrGlnGluArgLeuSerGlyAlaGlnAlaProAlaLeuTyrAlaIleSerLeuVal 881
||| ||| ||| |||
Qy 1286 ATCATCTTGCTCTATGTGAGTTTGGCGTTTGAATGCGCAATATGCTTTCAGTCCGTA--- 1342
||| ||| ||| ||| :||| |||
Db 882 ValPheLeuCysLeuAlaAlaLeuTyrGluSerTrpSerValProPheSerValMetLeu 901
||| ||| ||| ||| :||| |||
Qy 1343 -----TGGCTTTAATTATCATGACCTTTTGGCTACCTGTGCGAGTCTTGTATTATAGCA 1393
||| ||| ||| ||| |||
Db 902 ValValProLeuGlyValIleGlyAlaLeuLeuAlaThr----- 914
||| ||| ||| ||| |||
Qy 1394 CATTTCTTTTGAAGAAATTCAAATAGATTTCGAAGCCATTGGTGTCTTTAATGACTGTA 1453
||| ||| ||| ||| |||
Db 915 -----TrpMetArgGlyLeuGluAsnAspValTyrPheGlnValGlyLeuLeuThrVal 932
||| ||| ||| ||| :||| |||
Qy 1454 TTGGGTATTCAATTAACAATACCTTTGATCATTTTGTATCGTATTCGTGAAGATCGCCAA 1513
||| ||| ||| ||| :||| |||
Db 933 IleGlyLeuSerAlaLysAsnAlaIleLeuIleValGluPheAlaAsnGluMetAsnGln 952
||| ||| ||| ||| :||| |||
Qy 1514 GCG-----AACCTGTTT---ACCCTATGATGTTTGTAGTTAATGATGCCCTTCAAAAG 1564
||| ||| ||| ||| |||
Db 953 LysGlyHisAspLeuPheGluAlaThrLeuHis-----AlaCysArgGln 967
||| ||| ||| ||| :||| |||
Qy 1565 ACGTTCAGCCGCGTAAATGACAACA---GCTACAACCTATCATGATTTTGTAAATGCTT 1621
||| ||| ||| ||| |||
Db 968 ArgLeuArgProIleLeuMetThrSerLeuAlaPheIlePheGlyValLeuProMetAla 987
||| ||| ||| ||| :||| |||
Qy 1622 TTGTTTATAGCGGCTCTCTGCTCTTTAATTTTGCATTTATATGACCATAGGATCTTT 1681
||| ||| ||| ||| :||| |||
Db 988 ThrSerThrGlyAlaGlySerGlyGlyGlnHisAla-----ValGlyThrGlyValMet 1005
||| ||| ||| ||| :||| |||
Qy 1682 CTAGGAACCTTATCGTCTCTTTATATTGCA-----CCACCTCTCTGTTGTTTATG 1732
||| ||| ||| ||| :||| |||
Db 1006 GlyGlyMetIleSerAlaThrIleLeuAlaIleTyrPheValProLeuPheValLeu 1025
||| ||| ||| ||| :||| |||
Qy 1733 GTCCGTAATA 1741
||| ||| ||| |||
Db 1026 ValArgArg 1028
||| ||| ||| |||
```

Search completed: December 17, 2002, 12:36:29  
Job time : 73 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 17, 2002, 12:38:47 ; Search time 19.5 Seconds  
(without alignments)  
3185.498 Million cell updates/sec

Title: US-09-868-987-1

Perfect score: 3312

Sequence: 1 atggaactccgcatattgtc.....atgacaaattcagataatgc 1864

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 105981 seqs, 16662342 residues

Total number of hits satisfying chosen parameters: 211962

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09868987/runat 16122002 083009 23678/app query.fasta\_1.2055  
-DB=Published Applications AA -QFMT=faetan -SUFTX=rapb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09868987 @C@N 1 1 5 @tunat 16122002 083009 23678  
-NCPU=6 -ICPU=3 -NO XLPXY -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US08 NEW PUB.pcp.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT NEW PUB.pcp.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06 NEW PUB.pcp.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pcp.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07 NEW PUB.pcp.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pcp.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pcp.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW PUB.pcp.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pcp.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10 NEW PUB.pcp.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pcp.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW PUB.pcp.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2778	83.9	552	10	US-09-886-468-14
2	678.5	20.5	737	10	US-09-899-482-2
3	431	13.0	525	10	US-09-881-752A-96
4	364	11.0	323	10	US-09-899-482-4

Sequence 14, Appl  
Sequence 2, Appl  
Sequence 96, Appl  
Sequence 4, Appl

5 359.5 10.9 615 10 US-09-899-482-3  
6 277.5 8.4 637 10 US-09-852-053-3  
7 245.5 7.4 279 10 US-09-861-451A-28  
8 228 6.9 403 10 US-09-852-053-4  
9 147.5 4.5 1055 10 US-09-815-242-12189  
10 144 4.3 1034 10 US-09-815-242-10331  
11 136.5 4.1 1032 10 US-09-815-242-11145  
12 129 3.9 1040 10 US-09-812-020-282  
13 129 3.9 1040 10 US-09-815-242-10203  
14 128 3.9 1028 10 US-09-815-242-11516  
15 123.5 3.7 1025 10 US-09-912-020-283  
16 122.5 3.7 1062 10 US-09-815-242-5111  
17 122 3.7 761 9 US-09-895-913A-10  
18 122 3.7 1043 10 US-09-815-242-51194  
19 118 3.6 801 10 US-09-815-242-5330  
20 117.5 3.5 1049 10 US-09-815-242-10072  
21 116.5 3.5 1018 10 US-09-815-242-5186  
22 111.5 3.4 467 10 US-09-767-041-23  
23 111 3.5 597 12 US-10-139-876-20  
24 109 3.3 396 10 US-09-815-242-10968  
25 108 3.3 651 10 US-09-815-242-5331  
26 108 3.3 651 10 US-09-815-242-12682  
27 108 3.3 1049 10 US-09-815-242-13719  
28 106 3.2 488 12 US-10-094-059-4  
29 102.5 3.1 487 9 US-10-095-139-14  
30 102.5 3.1 487 10 US-09-795-693-27  
31 101 3.0 381 10 US-09-827-937A-18  
32 101 3.0 381 10 US-09-842-316-6  
33 101 3.0 381 10 US-09-731-030A-17  
34 99.5 3.0 881 10 US-09-982-736-2  
35 99.5 3.0 1053 10 US-09-815-242-5136  
36 98.5 3.0 396 10 US-09-741-669-395  
37 97.5 2.9 382 10 US-09-969-711-2  
38 96.5 2.9 341 9 US-09-971-228-13  
39 96.5 2.9 346 9 US-09-971-228-15  
40 96.5 2.9 362 10 US-09-993-844-8  
41 96.5 2.9 381 12 US-10-037-616-19  
42 96.5 2.9 382 9 US-09-971-228-5  
43 96 2.9 486 10 US-09-860-232A-7  
44 96 2.9 619 10 US-09-741-153-4  
45 94.5 2.9 314 9 US-09-971-228-14

#### ALIGNMENTS

#### RESULT 1

US-09-886-468-14  
; Sequence 14, Application US/09886468  
; Patent No. US20020037293A1  
; GENERAL INFORMATION:  
; APPLICANT: Aventis Pasteur Limited  
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses thereof  
; FILE REFERENCE: 77813-5  
; CURRENT APPLICATION NUMBER: US/09/886,468  
; CURRENT FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: 60/113,280  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,281  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,282  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,283  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,284  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,285  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,385  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/114,050  
; PRIOR FILING DATE: 1998-12-28  
; PRIOR APPLICATION NUMBER: 60/114,056  
; PRIOR FILING DATE: 1998-12-28

```

; PRIOR APPLICATION NUMBER: 60/114,057
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,058
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,059
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,061
; PRIOR FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 14
; LENGTH: 552
; TYPE: PRF
; ORGANISM: Chlamydia pneumoniae
; US-09-868-468-14

Alignment Scores:
Pred. No.: 1,57e-274 Length: 552
Score: 2773.00 Matches: 552
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 83.38% Indels: 0
DB: 10 Gaps: 0

US-09-868-987-1 (1-1864) x US-09-868-468-14 (1-552)
QY 101 ATGGTCAGCAGCCCTATTAAAGTCCCATTTGAAAAATCATGCCAGTGTCTCAGGAAA 160
DB 1 MetValSerSerProIleLeuAsnValProLeuYsAsnHisAlaSerValSerGlyLys 20
QY 161 TTTTACCACCGTGAAGTGAAGCAACCTGGCCCTCAGATTAAATCTGGAGCATCTCTTT 220
DB 21 PheThrIstArgGluValSerLysLeuAlaSerAspLeuYsSerGlyAlaMetSerPhe 40
QY 221 GTTCCCGAGGTTCTCAGTGAAGAGACGATCTCTCTGATCTTGGAAAAACAATGTACA 280
DB 41 ValProIleValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThr 60
QY 281 CAAGGCATTAATCTCAGCATGCTGTGGCTTGGCAATGCTTATTTGTTGATGAGCGTAT 340
DB 61 GlnGlyIleIleSerAlaCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80
QY 341 TATATGATTTGGAGCGCTATCTGCTTGGAGCTGTCTCTCGAATCTTTTGCTATCTGG 400
DB 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTyr 100
QY 401 GCGCTCTACAGTATTTGGATGGCCCACTGCTTGCAGACCTCGCTGGAGATTGTTCTT 460
DB 101 AlaIleLeuGlnIleValLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
QY 461 GCTATGGGATGGCCGTAGATGCAATGTTCTTGTATTCGAAGAATCCGAGAGAAATT 520
DB 121 AlaeMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluPhe 140
QY 521 TTATGTCTCAAGTCTTAAAAAATCTGTAGAAAAAGATATACCAAGCTTTTGGAGCC 580
DB 141 LeuLeuSerGlnSerIleuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
QY 581 ATTTTATTTACTGACTAGACGATATTTGGCTCAGACCTCTTTCTTCTGATATACA 640
DB 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180
QY 641 GGGCTTAAAGGGTGTGCTTGAATGATTTAGAAATTTTCTCTCAATGTTTACG 700
DB 181 GlyProIleuYsGlyPheAlaLeuThrLeuIleuGlyIlePheSerSerMetPheThr 200
QY 701 GCTCTTTTTCATGACTAAATTTTCTTCACTGTGTGATGAATTAAGCCCAACATACACAG 760
DB 201 AlaLeuPheMetThrLysPhePheMetLeuThrPheAsnLysTrpGlnHisThrGln 220
QY 761 TTGATATGATGATATGATGCTGGGGGTAAGCATGATTTCTTGAGAGATGAGAAAAA 820
DB 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLeuLys 240

```

```

QY 821 CTTTGGGCTGTTTCTGGAAGTGTCTTTTCTTTAGCTTGGCTGCTCGGGTTTGAAGCC 880
DB 241 LeuThrAlaValSerIleYsSerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260
QY 881 TGGATTTCCGTTTGGGAATGATTTTAAAGAGAGGATATGCTTTTACCTTTAATCCAAA 940
DB 261 TrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThrPheAsnProLys 280
QY 941 GAGCATGGCATGACGATGTGCTCAATGCGTGGCAAGTTGTGCATAAATACACAGAA 1000
DB 281 GlnHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300
QY 1001 GCTGTCTTTCTTCTGAGACTTCCCTATTCCAACCTTTGATCTTCAAGAAAAATCAAA 1060
DB 301 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLys 320
QY 1061 ATCTATTTAGTATTAAGCTTTAAGCTTAACTTAACTAGACATGACGCTCTCTCAAAA 1120
DB 321 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLys 340
QY 1121 TTACGATCATGAGCTGGCGTTATTTGGAGATTGTTGTCAAGAACAGGCTAGATTCTC 1180
DB 341 LeuThrIleMetSerThrArgTyrCysGlyIleValValArgAsnArgProArgPheLeu 360
QY 1181 TACGAAATCTTAAACGAAACGAAATTTTGTCTCAAGGTAAAGCAACATATGGAAG 1240
DB 361 TyrGlyAsnSerLysArgAsnAlaLysPheTrpSerLysValSerLysLeuSerLys 380
QY 1241 AAATGGGTTATCAGCGCATCGGCTTTTGAAGCTTTGGCAATCATCTGCTCTAT 1300
DB 381 LysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyr 400
QY 1301 GTGAGTTTGGCTTGAATGCAATATGCTTTTCAAGTCCGATATGCTTTAATTCATGAC 1360
DB 401 ValSerLeuArgPheGluThrGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAsp 420
QY 1361 CTTTGGCTTACCTGTGACGCTTGTATTATGCACTTTCTTTTGAAGAAAAATTCAAA 1420
DB 421 LeuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLysIleGlnIle 440
QY 1421 GATTGCAAGCCATGTGTGTTAATGACTGATTTGGGGATTCATTAACATACCTTTG 1480
DB 441 AspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnMetThrLeu 460
QY 1481 ATCATTTTGTATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1540
DB 461 IleIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThrPromethisVal 480
QY 1541 TTATGTTATGATGCTTCAAAAAGACGTTCAAGCCGCAAGGTAATGACACAGCTACACT 1600
DB 481 LeuValAsnAspAlaLeuGlnLysTrpPheSerArgThrValMetThrAlaThrThr 500
QY 1601 CTATCAGTTTGTATGCTTTGTTATAGGCGGCTCTGCTTATTTTATTTTGCATT 1660
DB 501 LeuSerValLeuLeuMetLeuLeuPheIleGlyGlySerSerValPheAsnPheAlaPhe 520
QY 1661 ATTATGACATAGGATTTCTTACGAACTTATTCGCTCTTATATTTGACACCACTCTG 1720
DB 521 IleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuThrIleAlaProPheLeu 540
QY 1721 TTGTTGTTATGCTCGGTAAAGAAATCGCTCAAA 1756
DB 541 LeuLeuPheMetValArgLysGluAsnArgSerLys 552

RESULT 2
US-09-899-482-2
; Sequence 2, Application US/09899482
; Patent No. US2002006641A1
; GENERAL INFORMATION:
; APPLICANT: Quak, Wilhelmus J.
; TITLE OF INVENTION: Increasing Production of Proteins in Microorganisms
; FILE REFERENCE: GC385-PC7

```



```

; CURRENT APPLICATION NUMBER: US/09/899,482
; CURRENT FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: EP 97305286.3
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: EP 97305344.0
; PRIOR FILING DATE: 1997-07-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Bacillus Subtilis
US-09-899-482-2

```

## Alignment Scores:

```

Pred. No.:      8,92e-61      Length:      737
Score:          678.50      Matches:      169
Percent Similarity: 50.43%      Conservative: 125
Best Local Similarity: 28.99%      Mismatches: 192
Query Match:      20.49%      Indels:      97
DB:              10          Gaps:      15

```

US-09-868-987-1 (1-1864) x US-09-899-482-2 (1-737)

```

Qy 104 GTCAGCAGCCCTATTAAACGTCCTCCATTGAAA-----AATCATGCCAGTGTCTCAGGGAAA 160
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 197 ValSerAlaProAsnValSerGlnGluLeuAsnThrThrAspValLysLeuGluGlyHis 216
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 161 TTTACCCACCGTGAAGTACGAACTCCCTCAGATTTAAATCTGGAGCGATGCTTTT 220
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 217 PheThrAlaGlnGluAlaLysAspLeuAlaSerIleLeuAsnAlaGlyAlaLeu----- 234
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 221 GTTCCCGAGGTTCTCAGTGAAGAG-----ACGATCTCTCTCTGATCTTTGGGAAAAA 271
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 235 ---ProValLysLeuThrGluLysTyrSerThrSerValGlyAlaGlnPheGlyGlnGln 253
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 272 CAATGTACAAAGCATATCTCAGCATCTGTGCTGGCTGGCAATGCTATTGTTTGATG 331
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 254 AlaLeuHisAspThrValPheAlaGlyIleValGlyIleAlaIleAlaIlePheLeuPheMet 273
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 332 AGCGTATATTATAGATTGGAGGCGTCATCGCTTCGGGAGCTGTTCTCTGAATCTTTTG 391
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 274 LeuPheTyrTyrArgLeuProGlyLeuIle-----AlaValIleThrLeuSerVal 290
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 392 CTTATCTGGCAGCTCTACAG-----TATTTCGATGGCCACTCACCTTGTTCAGGA 442
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 291 TyrIleTyrIleThrLeuGlnIlePheAspTrpMetAsnAlaValLeuThrLeuProGly 310
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 443 CTCGCTGGATTTCTTCGTATGGGATGGCCGTAGATGCAATGTTCTGTATTCGAA 502
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 311 IleAlaAlaLeuIleLeuGlyValGlyMetAlaValAspAlaAsnIleIleThrTyrGlu 330
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 503 AGAATCCGAGAGGAAATTTTATTGTCCTCAAGTCTTAAATAATCTGAGAAAAAGGATAT 562
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 331 ArgIleLysGluGluLeuLysLeuGlyLysSerValArgSerAlaPheArgSerGlyAsn 350
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 563 ACCAAGGCTTTGGAGCCATTTTTCATCTTAACCTTACACTACAGTATGCTCTCAGCACT 622
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 351 ArgArgSerPheAlaThrIlePheAspAlaAsnIleThrThrIleAlaAlaValVal 370
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 623 CTTTTCCTCCTAGATACAGGGCTATTAAAGGTTTGTCTTGGACATTTAGTAATTT 682
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 371 LeuPheIlePheGlyThrSerSerValLysGlyPheAlaThrMetLeuIleLeuSerIle 390
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 683 TTCCTCTCAATGTTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTG----- 733
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 391 LeuThrSerPheIleThrAlaValPheLeuSerArgPheLeuLeuAlaLeuValGlu 410
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 734 -----TGGATGAATAAG----- 745
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 411 SerArgTrpLeuAspArgLysLysGlyTrpPheGlyValAsnLysLysHisIleMetAsp 430
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 746 -----ACCAACAATACACAGTTGTCATATGATGAATAAGTTCTGCTGGGATAAAG 793

```

```

Db 431 IleGlnAspThrAspGluAsnThrGluProHisThrProPheGln-----Lys 446
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 794 CATGATTCTTTCAGAGGATGCAAAAACTTTGGCTGTTTTCGGAAGTGTTCCTTTT 853
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 447 TrpAspPheThrSerLysArgLysTyrPhePheIlePheSerSerAlaValThrValAla 466
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 854 GGTGCGCTTCTCTCGGGTTTGGAGCTCGGAATTCCTGTTGGGAATGGATTTAAAGGA 913
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 467 GlyIleIleIleLeuLeuValPheArgLeuAsn-----LeuGlyIleAspPheAlaSer 484
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 914 GGG-----TATGCGCTTTTACCTTTTAACTCCAAAAGACATGGC 949
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 485 GlyAlaArgIleGluValGlnSerAspHisLysLeuThrThrGluGlnValGluLysAsp 504
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 950 ATCAGCGATGTTGTCCTCAAAATCGCGCAAAATGTCATATAACTACAGGAAGCTGGTCTT 1009
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 505 PheGluSerLeuGlyMetAspProAspThrValValLeuSerGlyGlyLysSerAsnIle 524
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1010 TCT---TCTAGAGACTTCCGATTTCAAAACATTTGGATCTTCAGAAAAGATCAAAATCTAT 1066
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 525 GlyValAlaArgPheValGlyValProAspLysGluThrIleAlaLysValLysThrTyr 544
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1067 TTTAGTGTAAAGCTTTAAGCTATACTAAGCAGATACGAGCTCTCTCTTAAATTAACG 1126
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 545 PheLysAspLys----- 548
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1127 ATCATGAGCTGGCGTTATTGTTGGGATTGTTGTCAGAAAACAGCGCTAGATTTCCTACGGA 1186
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 549 -----TyrGly 550
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1187 AACTCTAAAGCAAGCAAAATTTTGTGTCAAAGTAAAGCAGCAAACTATCGAAGAAATG 1246
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 551 SerAspProAsnValSerThrValSerProThrValGlyLysGluLeuAlaArgAsnAla 570
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1247 CGTTATCAGCGACCATCGGGCTTTTAGGAGCTTTGGCAATCATCTTCCTCTATGTAGT 1306
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 571 LeuTyrAlaValAlaIleAlaSerIleGly-----IleIleIleTyrValSer 586
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1307 TTGGCTTTGAATGGCAATATGCTTTCAGTCCCTATGCGCTTTAATTCATGACCTTTTG 1366
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 587 IleArgPheGluTyrLysMetAlaIleAlaIleAlaIleAlaSerLeuLeuTyrAsp----- 604
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1367 GCTACCTGTGCAGTCTTCTTTATAGCACATTTCTTTTGAAGAAAAATTCAAATAGATTG 1426
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 605 -----AlaPhePheIleValThrPhePheSerIleThrArgLeuGluValAspVal 621
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1427 CAACCATGCTGCTTTAATGACCTGTTATGCGGTATTCATTAACAATACATCTTCATCAT 1486
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 622 ThrPheIleAlaAlaIleLeuThrIleIleGlyTyrSerIleAsnAspThrIleValThr 641
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1487 TTTGATCGTATTCCTGAA-----GATCGCAAGCGAACCTGTTTACCCCTATGCAT 1537
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 642 PheAspArgValArgGluHisMetLysLysArgLysProLysThrPheAlaAspLeuAsn 661
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1538 GTTTTAGTTAATGATGCCCTTCAAAAGACGTTCAAGCCGACGGTAATACACACAGCTACA 1597
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 662 HisIleValAsnLeuSerLeuGlnThrPheThrArgSerIleAsnThrValLeuThr 681
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1598 ACTCTACAGTTTGTAAAGCTTTTATAGCGCGCTCTCTGCTCTTAAATTTTCCA 1657
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 682 ValValIleValValThrLeuLeuIlePheGlyAlaSerSerIleThrAsnPheSer 701
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1658 TTTATTATGACCATAGGATTTCTTAGGAATTTTATCGTCTCTTATATGACCACCT 1717
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 702 IleAlaLeuValGlyLeuLeuThrGlyValTyrSerSerLeuTyrIleAlaAlaGln 721
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1718 CTGTTGTTG 1726
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 722 IleTrpLeu 724
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 3

US-09-881-752A-96

```
/ Sequence 96, Application US/09881752A
/ Patent No. US20020115078A1
/ GENERAL INFORMATION:
/ APPLICANT: Kleantous, Harold
/ APPLICANT: Al-Garawi, Amal
/ APPLICANT: Miller, Charles
/ APPLICANT: Tomb, Jean-Francois
/ APPLICANT: Oomen, Raymond P.
/ TITLE OF INVENTION: Identification of Polynucleotides
/ TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the
/ FILE REFERENCE: 06132/041002
/ CURRENT APPLICATION NUMBER: US/09/881,752A
/ CURRENT FILING DATE: 2001-06-15
/ PRIOR APPLICATION NUMBER: US 08/833,457
/ PRIOR FILING DATE: 1997-04-01
/ NUMBER OF SEQ ID NOS: 370
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 96
/ LENGTH: 525
/ TYPE: PRT
/ ORGANISM: Helicobacter pylori
/
US-09-881-752A-96

Alignment Scores:
Pred. No.: 1.2e-35 Length: 525
Score: 431.00 Matches: 99
Percent Similarity: 60.83% Conservative: 47
Best Local Similarity: 41.25% Mismatches: 92
Query Match: 13.01% Indels: 2
DB: 10 Gaps: 2

US-09-868-987-1 (1-1864) x US-09-881-752A-96 (1-525)
QY 47 GGACATATTTCT--GCAACCGTGATGCGCTATGCTGATGATGACGTTATATG 103
Db 274 GLyaspheSerGlyAlaSnValGlylsrghMetAlaIleValIleAspAsnlyVal 293
QY 104 GTCAGCAGCCCTATTATTAACGTCCTCATG--AAAAATCATGCCAGTGTCTCAGGAAA 160
Db 294 TySerAlaProValIleArgIluArgIleGlySerGlyGlnIleSerGlyAsn 313
QY 161 TTTACCAACCGTGAAGTGAAGCAATCGCCCTCAGATTAAATCGAGCGATGCTTTT 220
Db 314 PheSerValAlaGlnAlaSerAspLeuAlaIleAlaLeuAspSerGlyAlaMetSerAla 333
QY 221 GTTCCCGAGGTTCTCAGTGAAGAGACGATCTTCTGATCTTGAGGAAAAAACAATGACA 280
Db 334 ProIleGlnValIleuGlnlylsrghIleIleGlyProSerIleuGlylsrghVallys 353
QY 281 CAAGGCAATTATCTGACATGCTGTGGCTTGSCAAATGCTTATTTGTTGATGAGCGTATAT 340
Db 354 ThSerIleIleAlaIleuValIleGlyPheIleIleuValMetGlyPheMetValIleuTy 373
QY 341 TATGATTTGAGGCGCTCATGCTTGGAGCTGTTCTTCTGATCTTTTCTTATCTCGG 400
Db 374 TySerMetAlaGlyValIleAlaCysLeuAlaIleuValAlaIleuPheLeuIleVal 393
QY 401 GCACGCTACAGTATTTGATGCGCACTGCTTGTGACGATCGTGGGATTTGTTCTT 460
Db 394 AlaValMetAlaIlePheGlyAlaThreIleuThreIleuProGlyMetAlaGlyIleValIleu 413
QY 461 GCTATGGGATGCGCGTATATGATGCAAAATGTTCTTATTTGAAAGATCGAGAGAAATTT 520
Db 414 ThrValGlyIleAlaIleValAspAlaAsnIleIleIleAsnGlnlylsrghIleArgIluValIleu 433
QY 521 TTATGTCTCAAGCTTTAAAAATCTGTAGAAAAAGATATACCAAGCTTTTGAGACC 580
Db 434 ArgIluSnGlnlylsrghIleAlaIleAlaIleHisIleuGlnlylsrghIleAsnAlaSerArgAla 453
QY 581 ATTTTGTATCTTACCTGATGATGATGCTCAGACGATCTTTTCTTCTGATGATGACA 640
Db 454 IlePheAspSerAlaIleThreIleuIleAlaSerValIleuIleuTyAlaIleTyGlyThr 473
```

```
QY 641 GGGCTATTAAAGGTTTGCTTGTGACATTGATTAGAAATTTCTCTCAATGTTAGC 700
Db 474 GLyAlaIlelySerGlyPheAlaIleuThrGlnlyIleGlyIleuAlaSerIleIleThr 493
QY 701 GCTCTTTTCATGACTTAAATTTTCTTCATGCTGTGATGATTAAGCCAAATACAGAG 760
Db 494 AlaIleValGlyThrGlnGlnlyIleTyGlnAlaIleuLeuProlylsrghIleuThrGlnThrly 513

RESULT 4
US-09-899-482-4
/ Sequence 4, Application US/09899482
/ Patent No. US20020006641A1
/ GENERAL INFORMATION:
/ APPLICANT: Quax, Wilhelmus J.
/ TITLE OF INVENTION: Increasing Production of Proteins in Microorganisms
/ FILE REFERENCE: GC385-PCT
/ CURRENT APPLICATION NUMBER: US/09/899,482
/ CURRENT FILING DATE: 2001-07-05
/ PRIOR APPLICATION NUMBER: EP 97305286.3
/ PRIOR FILING DATE: 1997-07-16
/ PRIOR APPLICATION NUMBER: EP 97305344.0
/ PRIOR FILING DATE: 1997-07-17
/ NUMBER OF SEQ ID NOS: 4
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 4
/ LENGTH: 323
/ TYPE: PRT
/ ORGANISM: E. coli
/
US-09-899-482-4

Alignment Scores:
Pred. No.: 6.47e-29 Length: 323
Score: 364.00 Matches: 102
Percent Similarity: 49.52% Conservative: 53
Best Local Similarity: 32.59% Mismatches: 124
Query Match: 10.99% Indels: 34
DB: 10 Gaps: 9

US-09-868-987-1 (1-1864) x US-09-899-482-4 (1-323)
QY 794 CATGATTTCTTGAGAGATGCAAAAACTTGGGCTGTTCTGGAAGTGTCTTTCTTTTA 853
Db 17 TyAspPheMetArgIleThrAspIlyrTrpAlaPheGlyIleSerGlyIleuLeuIleAla 36
QY 854 GGTGGCTGCTCTCGGCTTTGAGCGTGGAAATTCGCTTTGGAAATGATTTTAAAGA 913
Db 37 AlaIleValIleMetGlyValArgGlyPheAsn-----TrpGlyLeuAspPheThrGly 54
QY 914 GGGTATGCTTACCTTTATCCAAAAGACATGCAATGATGCTGCTCAAAATGCGT 973
Db 55 GlyThrValIleGlnIleThreIleuGlnlylsrghIleAsnIleAspVal-----MetArg 72
QY 974 GGCAAAGTTGTGATTAACATGACGAGGCTGCTTCTTCTGAGACCTTCGATTTCAA 1033
Db 73 -----AspAlaIleuGlnlylsrghIlePhe-----GluIleuProMetLeuGln 86
QY 1034 ACATTTGATCTTCAAGAAAAGATCAAAATCTATTTTATGATTAACCTTTAAGCTTACT 1093
Db 87 AsnPheGlySerSerHisAspIleMetValArgMetProIleAlaGlnlylsrghIle 106
QY 1094 AAGCATATGACGCTCT--CTCTTAAATTTAAGCATGATGCTGCGCTTATTTGGG 1150
Db 107 GlyGlnValIleuGlylsrghIleValIleuTyValIleAsnIleuSer----- 121
QY 1151 ATGTTGTGAGAAACAGGCTTATGATTTCTCTAGAAACTCTTAAACGAAACGAAATTT 1210
Db 122 -----ThrsnGlnAsnAlaIleValIleValys 130
QY 1211 TGGTCAAGGTAGACAGCAAACTATCGAAGAAATCGTTATCGAGCGACATCGGCGTT 1270
Db 131 IleGlnPheValIleProSerValIleGlyAlaAspLeuAlaGlnThrGlyAlaMetAlaIleu 150
```

```
QY 1271 TTAGGAGCTTTGGCAATCATCTTGCTCTATGTGAGTTTGGCTTTGAATGCAATATGCT 1330
Db 151 MetAlaLeuLeuSerIleLeuValTyrValGlyPheArgPheGluTrpArgLeuAla 170
QY 1331 TTCAGTCCGCTATGCGCTTAATTCATGACCTTTGGCTGACCTGCGCAGTCTTGTTATA 1390
Db 171 AlaGlyValValIleAlaLeuAlaHisaspValIleIleThrLeuGlyIleLeuSerLeu 190
QY 1391 GCACATTTCTTTTGAAGAAATCAATAGATTTGCAAGCATTGGTGGCTTTAATGACT 1450
Db 191 PheHis-----IleGluIleAspLeuThrIleValAlaSerLeuMetSer 205
QY 1451 GTATTGGGATTTCAATAAATCTTATGATCTTTTGTATGCTGATTCGTCAGAT--- 1507
Db 206 ValIleGlyTyrSerLeuAsnAspSerIleValValSerAspArgIleArgGluAsnPhe 225
QY 1508 GCCCAAGCGAACTCTTTACCCCTATGATGTTTGTAGTTAATGATGCCCTTCMAAAGACG 1567
Db 226 ArgIleValIleArgGlyThrProTyrGluIlePhe---AsnValSerLeuThrGlnThr 244
QY 1568 TTCAGCCGACGGTAATGACACAGCTCAACTCTATCAGTTTGTGTTTGTGTTT 1627
Db 245 LeuHisArgThrLeuIleThrSerGlyThrThrLeuMetValIleLeuMetLeuTyrLeu 264
QY 1628 ATAGCGGCTCTCTGCTCTTTAATTTGATTTATGACCATAGGATTTCTTAGCA 1687
Db 265 PheGlyGlyProValLeuGluGlyPheSerLeuThrMetLeuIleGlyValSerIleGly 284
QY 1688 ACTTTATCGTCTCTTTATATTGCACCACTCTGTTGTTG 1726
Db 285 ThrAlaSerSerIleTyrValAlaSerAlaLeuAlaLeu 297
RESULT 5
US-09-899-482-3
; Sequence 3, Application US/09899482
; Patent No. US20020006641A1
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: Increasing Production of Proteins in Microorganisms
; FILE REFERENCE: GC385-PCT
; CURRENT APPLICATION NUMBER: US/09/899,482
; CURRENT FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: EP 97305286.3
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: EP 97305344.0
; PRIOR FILING DATE: 1997-07-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 615
; TYPE: PRT
; ORGANISM: E. coli
US-09-899-482-3
Alignment Scores:
Pred. No.: 2.44e-28 Length: 615
Score: 359.50 Matches: 80
Percent Similarity: 62.39% Conservative: 61
Best Local Similarity: 35.40% Mismatches: 82
Query Match: 10.85% Indels: 3
DB: 10 Gaps: 2
US-09-868-987-1 (1-1864) x US-09-899-482-3 (1-615)
QY 44 AATGGCAATATCTTCAACCGTGTAGTGGCTGTAGTCAATGACGGTTATATG 103
Db 375 SerGlyLysLysAspAlaAsn-----GlyArgAlaValLeuValIleGlnGluVal 392
QY 104 GTCACGACCGCTTATTAAACGCTCCCATGAAATCATGCCAGTGCTCAGGGAATTT 163
Db 393 IleAsnIleAlaAsnIleGlnSerArgLeuGlyAsnSerPheArgIleThrGlyIleAsn 412
QY 164 ACCCAACCGTGAAGTAGGCAAACTCGCCTCAGATTTAAATCTGGAGCGCATGCTTTGTT 223
```

```
Db 413 AsnProAsnGluAlaArgGlnLeuSerLeuLeuLeuArgAlaGlyAlaLeuIleAlaPro 432
QY 224 CCCAGAGTTCTCAGTGAAGACGACGATCTCTCTCATCTTGGGAAAAAACAATGTACACAA 283
Db 433 IleGlnIleValGluAlaArgThrIleGlyProThrLeuGlyMetGlnAsnIleGluGln 452
QY 284 GGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTATTTGTTGATGAGGCTATATAT 343
Db 453 GlyLeuGluAlaCysLeuAlaGlyLeuValSerIleLeuPheMetIleIlePheTyr 472
QY 344 AGATTGGAGCGCATCGCTTCCGGAGCTGTTCTCTCTGAATCTTTGCTTATCTGGCA 403
Db 473 LysLysPheGlyLeuIleAlaThrSerAlaLeuIleAlaAsnLeuIleLeuValGly 492
QY 404 GCTCTACAGTATTG---GATCGCCCACTCACCTTGTCCAGACTCGCTGGGATTTGTTCTT 460
Db 493 IleMetSerLeuLeuProGlyAlaThrLeuSerMetProGlyIleAlaGlyIleValLeu 512
QY 461 GCTATGGGATGCGCGTAGATGCAATGTTCTTGTATTCGAAAGAAATCCGAGAGAAATTT 520
Db 513 ThrLeuAlaValAlaValAspAlaAsnValLeuIleAsnGluArgIleLysGluGluLeu 532
QY 521 TTATTGCTCAAGCTCTTAAAAAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCC 580
Db 533 SerAsnGlyArgThrValGlnAlaIleAspGluGlyTyrArgGlyAlaPheSerSer 552
QY 581 ATTTTTCATCTCACTTACACTACAGTATTGGCTTCAGCACTTCTTTCTCTCTAGATACA 640
Db 553 IlePheAspAlaAsnIleThrThrLeuIleValIleLeuIleLeuTyrAlaValGlyThr 572
QY 641 GGCCTATTAAAGGTTTGTCTTGCATGATGATTTTGAAGAAATTTCTCTCAATGTTTACG 700
Db 573 GlyAlaIleLysGlyPheAlaIleThrThrGlyIleGlyValAlaThrSerMetPheThr 592
QY 701 GCTCTTTTCATGACTAAA 718
Db 593 AlaIleValGlyThrArg 598
RESULT 6
US-09-852-053-3
; Sequence 3, Application US/09852053
; Patent No. US20020055141A1
; GENERAL INFORMATION:
; APPLICANT: BERENS, STEPHAN
; APPLICANT: KALINOWSKI, JORN
; APPLICANT: PUHLER, ALFRED
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM STRAIN WITH
; ENHANCED SECRETION ACTIVITY
; FILE REFERENCE: WAS/21123/280248
; CURRENT APPLICATION NUMBER: US/09/852,053
; CURRENT FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: EPO 00110021.3
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: PROPEP
; LOCATION: (1)..(637)
; OTHER INFORMATION: secD
US-09-852-053-3
Alignment Scores:
Pred. No.: 5.5e-20 Length: 637
Score: 277.50 Matches: 67
Percent Similarity: 51.81% Conservative: 62
Best Local Similarity: 26.91% Mismatches: 87
Query Match: 8.38% Indels: 33
DB: 10 Gaps: 7
```

US-09-868-987-1 (1-1864) x US-09-852-053-3 (1-637)

QY 5 ACTTCCCATATTTGTCAGAGGGATGACGGCAGCTGCTAATGACAATATTCGAAAC 64  
 Db 326 ThserGlnTyrLeuGlnGln----- 332  
 QY 65 CGTGGATGGCGTATGCGTATGATGAGCGTTATATGTCACAGCCCTATTTTAAAC 124  
 Db 333 -----GlnIleAlaIleThrLeuAspSerGlnValIleSerAlaProValIleGln 349  
 QY 125 -----GTCCGATTGAAAATCATGCGCAGTGTCTGAGGAAATTTTACCCACCGTGAAGTG 178  
 Db 350 SerAlaThrProValGlySerAlaThrSerIleThrGlyAspPheThrGlnIleGlnAla 369  
 QY 179 AGCAAACTCGCCTCAGATTAAATCTGAGCG-----ATGCTTTTGTTCGCGAG-- 229  
 Db 370 GlnAspLeuAlaAsnIleuAlaGlyTyrGlyAlaLeuProLeuSerPheAlaGlyGlnAsn 389  
 QY 230 -----GTTCTCAGTGAAGAGACGATCTCTTCTGATCTTGGGAAAAACAATGAC 280  
 Db 390 GlyGlnArgGlyGlyThrThrThrValProProSerLeuGlyAlaAlaSerLeuLys 409  
 QY 281 CAAGGCAATATCTCAGATGCTGTGGCAATGCTTATGTTGATGAGCGTATAT 340  
 Db 410 AlaGlyLeuIleAlaGlyIleValGlyIleAlaLeuValAlaIlePheValPheAlaTyr 429  
 QY 341 TATAGATTGGAGCGCTGATC-----GCTCGGAGCTGTTCTCTG 382  
 Db 430 TyrTrpValPheGlyPheValSerLeuPheThrLeuPheAlaIleGlyValIleValTyr 449  
 QY 383 AATCTTTGCTTATCTGAGCGCTCTACAGATTTTGGATGGCGCACTACCTTGTGAGA 442  
 Db 450 GlyLeuLeuValIleLeuGly-----ArgTrpIleGlyTyrSerLeuAspLeuAlaGly 467  
 QY 443 CTCGCTGGGATGTTCTTCTGATAGGATGCGCCGAGATGCAAAATGCTTGTGATGCA 502  
 Db 468 IleAlaGlyLeuIleIleGlyIleIleIleThrAlaAspSerPheValValPheTyrGln 487  
 QY 503 AGAATCCGAGAGAAATTTTATGTTCTCMAAGCTTAAATAATCTGTGAAAAAGATAT 562  
 Db 488 ArgIleLysAspGluIleArgGlnGlyArgSerPheArgSerAlaValProArgAlaTrp 507  
 QY 563 ACCAAGCTTTGAGCCATTTTGTATCTTAACTTGAAGTATGAGTTGGCTGACACTT 622  
 Db 508 GluSerAlaLysArgThrIleValThrGlyAsnMetValThrLeuGlyAlaIleVal 527  
 QY 623 CTTTCTTCCATGATGACGAGCGCTATTAAGGTTTGGCTTGAAGTATG 679  
 Db 528 IleTyrLeuLeuAlaValGlyGlnValIleGlyPheAlaPheThrLeuGlyLeuThr 547  
 QY 680 ATTTTCTCTCATGTTTACGCTCTT 706  
 Db 548 ValPheAspLeuValValThrPheLeu 556

RESULT 7  
 US-09-861-451A-28  
 ; Sequence 28, Application US/09861451A  
 ; Patent No. US2002006829A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Commonwealth Scientific & Industrial Research Orga  
 ; TITLE OF INVENTION: Methods of Identifying Antigen Gene Sequences  
 ; FILE REFERENCE: FP34033/01  
 ; CURRENT APPLICATION NUMBER: US/09/861,451A  
 ; PRIORITY FILING DATE: 2001-05-21  
 ; PRIOR APPLICATION NUMBER: PP2273  
 ; NUMBER OF SEQ ID NOS: 84  
 ; SOFTWARE: Patent Ver. 2.0  
 ; SEQ ID NO 28  
 ; LENGTH: 279  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence

FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Deduced protein  
 ; OTHER INFORMATION: Sequence from clone PAD727  
 US-09-861-451A-28

## Alignment Scores:

Pred. No.:	7,02e-17	Length:	279
Score:	245.50	Matches:	77
Percent Similarity:	49.50%	Conservative:	72
Best Local Similarity:	25.41%	Mismatches:	105
Query Match:	7.41%	Indels:	49
DB:	10	Gaps:	9

US-09-868-987-1 (1-1864) x US-09-861-451A-28 (1-279)

QY 893 TTGGGAATGATTTTAAAGAGGATAGCTTATCTTAAATCAAAAGATGCGATC 952  
 Db 8 LeuAlaIleAspPheSerGlyThrAsnPheLeuIleGluSerSerAsnSerTyr 27  
 QY 953 AGCGATGCTCTCAATGCGTGCAGAAAGTTGTCATTAACATGACGAGAGCTGTTCT 1012  
 Db 27 AspLeuIleThrLysGlnLysAlaGlyLysIleIleSerPhe----- 41  
 QY 1013 TCTAGAGACTTCGATTTCAACATTTGATCTTCAGAAAAGATCAAAATCTATTAGT 1072  
 Db 42 -----LeuAspSerGlnAsnIleAsnLysSerAsnSerThr 53  
 QY 1073 GATTAAGCTTTAAGCTTAAAGACAGATACAGAGCTCTCTCAAAATTAAGATCATG 1132  
 Db 53 IleLeuLeu-----AsnProLeuAsnGlnAsn----- 62  
 QY 1133 AGCTGGGTTATTTGTGAGATTTGTCACAAAAGAGCTTGAATTTCTTACGGA----- 1187  
 Db 63 -----GlyAsnIlePheAsnLeuIleLysTrpLysLeuAspLeuAlaThrLysIleAl 81  
 QY 1188 -ACTCTTAAAGAAAGC-----CAAAATTTTGTGTC-----AAGT 1221  
 Db 81 AserLeuAsnThrAlaIleGlnAsnAsnPheSerAsnIleArgMetThrAsnTyrSerIle 101  
 QY 1222 AAGCAGCAAACTATGCAAGAAATGCGTTATACGCGACCATCGGCTTTAGAGCTTT 1281  
 Db 101 eSerAsnGlnGluAlaGlnLysLeuIlePheAsnAlaIleLeuSerValGlyIleAlaIle 121  
 QY 1282 GGCATCATCTGCTCATGTGAGTTGGCTTTGAATGGCAATATGCTTACGTCCG 1341  
 Db 121 uIlePheValThrIlePheThrLeuIleArgPheLysTrpThrPheSerLeuAlaIleI 141  
 QY 1342 ATGCGCTTATATCATGACCTTTGCTACCTGTGACAGTCTGTTATATGACATTTCTT 1401  
 Db 141 ePheSerLeuLeuPheAsnValLeuMetValIleLeuAlaIleIleThr----- 158  
 QY 1402 TTGAAGAAAATTCAAATAGATTGCAAGCATGTGCTTATATGACTGTATGGGATA 1461  
 Db 159 -----ArgIleGlnIleSerGlnAsnLeuValValAlaIleLeuThrLeuIleGly 176  
 QY 1462 TTCAATTAACAATATCTTATGATCATTTTGAATGCTATTCG-----GAAGATCG 1509  
 Db 176 ThrValAsnAspThrIleValValPheAspArgValLysAlaArgPheSerLysLeuAs 196  
 QY 1510 CCAAGCGAACCTG-----TTTACCCCTATGATGATTTTGTAAATGATCCCTCAAAA 1563  
 Db 196 nHISGlnAsnValTyrLysPheAspLysIleLysGlnIleSerLeuGlnAlaIleArgI 216  
 QY 1564 GAGCTTCAGCGGACGTAATGACAACAGCTACACTCATATGATTTTGAATGCTTTT 1623  
 Db 216 uThrAlaLysArgSerValTyrThrSerLeuThrIleLeuThrIleValValLeuMe 236  
 QY 1624 GTTTATAGCGGCTCTCTGCTTATTTTGAATTTATATATACCATATAGGATTTCTT 1683  
 Db 236 tIlePheTyrGlnSerIleAspIleValPheSerLeuThrMetLeuIleGlyValIleI 256  
 QY 1684 AGGAACTTATGCTCTTATATGCAACCACTGTTGTTATGTCGCTAAGA 1743  
 Db 1684 ----- 1743

Db 256 eGlyThrTyrSerSerLeuPheAlaThrArgIleTrpIleLeuLeuGluSerSerAr 276  
Qy 1744 AAATCGC 1750  
Db 276 gAsnArg 278  
RESULT 8  
US-09-852-053-4  
; Sequence 4, Application US/09852053  
; Patent No. US20020055141A1  
; GENERAL INFORMATION:  
; APPLICANT: BERENS, STEPHAN  
; APPLICANT: KALINOWSKI, JORN  
; APPLICANT: PUHLER, ALFRED  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM STRAIN WITH  
; TITLE OF INVENTION: ENHANCED SECRETION ACTIVITY  
; FILE REFERENCE: MAS/21123/280248  
; CURRENT APPLICATION NUMBER: US/09/852,053  
; CURRENT FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: EPO 00110021.3  
; PRIOR FILING DATE: 2000-05-12  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 403  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
; FEATURE:  
; NAME/KEY: PROPEP  
; LOCATION: (1)..(403)  
; OTHER INFORMATION: secf  
US-09-852-053-4  
Alignment Scores:  
Pred. No.: 4,95e-15 Length: 403  
Score: 228.00 Matches: 72  
Percent Similarity: 44.97% Conservative: 80  
Best Local Similarity: 21.30% Mismatches: 132  
Query Match: 6.88% Indels: 54  
DB: 10 Gaps: 8  
US-09-868-987-1 (1-1864) x US-09-852-053-4 (1-403)  
Qy 797 GATTCTTGTAGAGGATGCAAAATCTTGGCTGTTCGTGAAGTGTTTTCTTTTAGGT 856  
Db 34 AspPheIleAlaLysThrLysLeuTrpTyrTrpIleThrGlyIleLeuLeuValIleSer 53  
Qy 857 TCGGTGCTCTCGGTTTGGAGCCTGGAATCCGTTTGGGATGATTTAAAGGAGGG 916  
Db 54 IleLeuPheIleAlaLysThrLysLeuTrpTyrTrpIleThrGlyIleLeuLeuValIleSer 71  
Qy 917 TATGCTTTTACCTTTAATCCAAAGAGCATGGCATCGATGTTGCTCAAAATGCGTGGC 976  
Db 72 ThrLysMetSerMetProAlaSerAspTyrSerThrGluInValGluGluThr----- 89  
Qy 977 AAAGTTGTGCATAAACTACAGAAAGCTGTCTTTCTTAGAGACTTCGGTATTCAACA 1036  
Db 90 -----PheThrGluAlaThrGlyIleThrProGluIleValGlnIleValGly 105  
Qy 1037 TTTGGATCTTCAGAAAGATCAAAATCTATTATTAGTATAAGCTTTAAAGCTATACTAAG 1096  
Db 106 SerGlyAspAlaArgThrLeuGluIleTyr-----SerGluArgLeuSerAspGluAsp 123  
Qy 1097 CAGATACAGAGCTCTCTCTCTAAATTAACGATCATGAGCTGCGTTATTGTGGGATGTT 1156  
Db 124 ValGluLysAlaArgLeu-----AlaIleTyr 132  
Qy 1157 GTCAGAAACAGCCTAGATTTCTTACGGAACCTCAAAACGAAACGCAAAATTTTGTCA 1216  
Db 133 GluGluTyrGlnProLeuAsnSerGluGlnProSerProAspAlaIleGlyAsnSer 152  
Qy 1217 AAGTTAACGACGCAAACTATCGAGAAATGCGTTATCAGGGACCATCGGCTTTTGA 1276

Db 153 ThrValSerGluSerTrpGlySerThrIleThrGlnArgMetValLeuAlaLeuIleAla 172  
Qy 1277 GCTTTGGCAATCATCTTCTCTATGTAGTTTGGCTTTTGAATGGCAATATGCTTTTCA 1336  
Db 173 PheLeuValIleAlaAlaIleTyrIleAlaPheArgLeuGluArgGluMetAlaIleAla 192  
Qy 1337 GCGTATGCGCTTTAATTCATGACCTTTTGCTACCTGTGTGCAGCTCTGTTTATAGC 1396  
Db 193 AlaMetAlaAlaLeuValValAsp-----GlyIleValIleAlaGlyIle 207  
Qy 1397 TCTTTTGTGAAGAAAATTCAAATAGATTGTCAGCCATTGCTGCTTTAATGACTGTATTG 1456  
Db 208 TyrAlaValIleGlyLeuGluValSerProAlaThrValIleGlyLeuLeuThrVal 227  
Qy 1457 GGGTATTCAATTAACAATCTTTCATCATCTTTTCATCGTATTTCGTGAAGAT----- 1507  
Db 228 ThrPheSerIleTyrAspThrValValValPheAspLysValArgGluAsnThrGluGly 247  
Qy 1508 -----CGCCAAGGGAACCTGTTTACCCTATGTCAT 1537  
Db 248 PheGluGlySerArgArgThrTyrAlaGluGlnAlaAsnLeu----- 262  
Qy 1538 GTTTTAGTTAATGATGCTTCAAAAGACGTTTCAGCGCAGGTAATGACACAGACTACA 1597  
Db 263 -----AlaValAsnGlnThrPheMetArgSerIleSerThrIleIle 277  
Qy 1598 ACTCTATCAGTTTGTGTTAATGCTTTGTTTATA-----GGCGGCTCTCT 1642  
Db 278 SerAlaLeuProIleIleAlaLeuMetValValAlaValTyrMetMetGlyValGlyThr 297  
Qy 1643 GTCCTTAAATTTTGGATTTATATGACCATGGGATTTCTTAGGAATTTATGCTCTCTT 1702  
Db 298 LeuLysAspLeuAlaLeuIleGlnLeuIleGlyValIleGluGlyThrPheSerSerVal 317  
Qy 1703 TATATTGACCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1756  
Db 318 PheLeuAlaThrProLeuLeuValSerLeuLysAsnArgLeuSerLysThrLys 335  
RESULT 9  
US-09-815-242-12189  
; Sequence 12189, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: EUTRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12189

```

; LENGTH: 1055
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12189

Alignment Scores:
  pred. No.:      1,16e-06      length:      1055
Score:           147.50        Matches:      132
Percent Similarity: 34.42%      Conservative: 111
Best Local Similarity: 18.70%      Mismatches: 229
Query Match:      4.45%         Indels:      234
DB:               10           Gaps:      28

US-09-868-987-1 (1-1864) x US-09-815-242-12189 (1-1055)

QY 265 GAAAAAACAATGTACACAGGCAATTATCTACAGTCTGTGGCTTGGCAATGCTTAATGT 324
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 366 GluyserserleuYrThnMet-ValGluYsAlaserleuYrThrleValAlaIleI 385
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 325 TTTCATGAGCGTATATTATAGA-----TTTGAGGCGTCAATGCTTCGGAGCTGTCT 378
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 385 evalIleleuYrThnMetArgGlnIleValSerleIleSerleIleSerleIlePr 405
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 379 TCTGAATCTTTTGTCTATCTGAGCGCTCTACAGTATTGATGCGCACTGACCTTGTCT 438
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 405 oleuSerleuYrThnMetAlaIleValleuYsAlaserleuYrThrleValAlaIle 425
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 439 AGGACGCTGGGATGCTTCTGTATGGGATGCGCGTACAGTCAATGCTTGTAT 498
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 425 uThrleuYrThnMetAlaIleValleuYsAlaserleuYrThrleValAlaIle 445
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 499 CGAAGATC-----CGAGAGATTTTATTGCTCAAACTCTTAA 540
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 445 lGluAsnIleYrThnMetArgGlnIleValSerleIleSerleIleSerleIlePr 464
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 541 AAATCTGTGAAAAAGATATACCAAGGCTTTGGAGCCCTTTTGTATCTTACCTGAC 600
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 465 -----IleIleSerleIleValleuYsAlaserleuYrThrleValAlaIle 482
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 601 TACAGTATGGCTCAGCACTTCTTCTCTCTATACAGGCGCTAT 649
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 482 lThrIleIleValleuYrThnMetArgGlnIleValSerleIleSerleIleSerle 501
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 650 -AAAGGCTTGTGACATTTAGCAATTTCTCTTCAATGCTTACG----- 700
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 501 eArgProHeuAlaIleValleuYsAlaserleuYrThrleValAlaIle 521
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 701 -----GCTCTTTCATGATTAATTTCTTCATGCTGAGTGAATAGACCA 750
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 521 rLeuValProAlaIleValleuYsAlaserleuYrThrleValAlaIle 541
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 751 ACATACAG-----TTGCATATGATG- 772
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 541 nHISGlnGluYrThnMetArgGlnIleValSerleIleSerleIleSerleIle 561
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 772 ----- 772
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 561 uAsnHISGlnGluYrThnMetArgGlnIleValSerleIleSerleIleSerleIle 581
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 773 -----AATAGTTCGAGGAT 789
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 581 eGluYrThnMetArgGlnIleValSerleIleSerleIleSerleIleSerleIle 601
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 790 AAACGAT----- 796
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 601 eThrYrThnMetArgGlnIleValSerleIleSerleIleSerleIleSerleIle 621
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 797 -----GATTTCTGAGGATGCAAACTTGGGCTGT 832
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 621 lGluYrThnMetArgGlnIleValSerleIleSerleIleSerleIleSerleIle 641
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 833 -----TCGGAAGTGTCTTTCATGCTTGTCT 865
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 641 rSerProValAspProThnGlySerThrAsnSerMetAlaIleMetValGluYrAspAs 661
QY 865 ----- 865
Db 661 nAspThrProAsnPheAspValGluAlaAspValIleYsHISAlaAspGlyPheY 681
QY 866 -----CTCGGCTTGGAGCTGGAATTCGTTTGGG 897
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 681 nHISProGluYrThnMetArgGlnIleValSerleIleSerleIleSerleIle 701
QY 898 AATGATTTTAAAGAGGATGCTTTCATTTAAATCCAAAGACATGCGATGACGCA 957
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 701 uValThrValYsGlyPro-----SerMetAspAlaIleYrSerThrValYsAs 718
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 958 TGTTCCTCAATTCGCGCAAGTT----- 982
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 718 pIleGlnGlnYsMetLeuGlnValYsGlyLeuAlaAsnValYsSerAspSerleu 738
QY 983 -----GTGCAATAACTACAGGAAGCTGCTTTC 1011
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 738 nHrYrAspGlnYrGluIleYsValAspGlnAsnYsAlaAlaGluAsnYsIleSe 758
QY 1012 TCTACAGACTTCCTGAT 1030
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 758 rAlaSerGlnLeuValMetHISleuAsnGluAsnLeuProGluYrThrValIleThr 778
QY 1031 -----CAACATTTGGATC 1044
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 778 lYsGluAsnGlyLeuThrValGlyValYsValYsGlnAsnYsGlnThrAspTrpSe 798
QY 1045 TTCGAAAGATCAAAATCATTTTGTGATTAAGCTTAAGCTATACAGAGATAG 1104
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 798 rGluAspYsLeuAsnAsnIleThrleuYsYsProThnGlyYrThrIleuYsGlu 818
QY 1105 A-----GCCTCTCTCTAAATTAACGATCATGAC-----TGGCG 1140
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 818 yAspIleAlaThrleuValYsThrThrProSerYsleuThrGlnGluGlnYsAs 838
QY 1141 TTAATGCGATTTGTT-----GTCAAGAACAGCGCTAGATTCTTACGAAATC 1191
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 838 pTYrAlaThrThrValSerAlaYsValThrAsnYsAsp-----ValGlyGlyTh 855
QY 1192 TAAACGAAAGCAAAATTTGTCGAAAGTACACCAAACTGACAGAAATGCGTTA 1251
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 855 rHnArg-----GlnAlaMetSerYsIleAsnAsn-----LeuAspYrProAsnAsn 872
QY 1252 TCAGGCGACCATCGCG----- 1267
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 872 lYsValAsnIleGlyGlyAlaSerAspAspIleAsnAsnAlaMetThrGlnleuAla 892
QY 1268 -----CTTTAGGAGCTTTGGCAATCATCTTGCTATGAGTTCGCTTGAATGCA 1323
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 892 eAlaMetleuAlaIleIleIleValIYrleuIleleuValIleThrPheYsGly 912
QY 1324 ATATGCT-----TTGAGTCCGCTATGCGCTTAATCATGACCTTTGGCTACGCTGAG 1380
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 912 yLeuAlaProPheThrIleuPheSerleuProPheThrValIleGlyValIleIle 932
QY 1381 CTGTTTATAGACATTTCTTTTGAAGAAATTAATGATGATTTGCAAGCCATGGTGC 1440
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 932 aleuIleIleThrGly-----GluThrIleSerValPro-----SerleuIleGly 948
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1441 TTTAATGATGTAATGGGATTAATTAACAAATCTTGATCATTTTGGATGTAATCG 1500
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 948 tLeuMet-----LeuIleGlyLeuValThrAsnAlaIleValleuIleAspYrVal 967
QY 1501 TGAAGATCGGCAAGCAACCTGTTACCTTACCGATGTTTGTAGTAAATGACCCCTTCA 1560
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 967 eAsnAsnGlnGlnGlnGlyMet-----GluMetYsGluAlaLeuIleGluAlaGly 985
QY 1561 AAAGACGCTGACCGGCAACGTAATGACAAACAGCTAATCTTCACTTTGTTAAATGCT 1620
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 985 yThnArgIleArgProIleuMetThrAlaIleAlaThrIleGlyAlaLeuValProle 1005
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Qy 1621 TTTGTTTATAGCGCTCTCTGCTCTT-----AATTTTGCAATTATTATGACCAT 1671
|||||
Db 1005 uLeuPheGlyGlnAspSerIleLeuIleSerIysGlyLeuAlaAlaThrValIleG1 1025
|||||
Qy 1672 AGGATTCTTCTAGGAATTTATGCTCTCTTATATTGACCACTCTG-----TTGTT 1725
|||||
Db 1025 yGlyLeuIleSerThrLeuLeuThrLeuValValProValIleTyrGluIleLe 1045
|||||
Qy 1726 GTTATGTCCTGATAA 1741
|||||
Db 1045 uPheThrLeuLysLys 1050
|||||

RESULT 10
US-09-815-242-10331
; Sequence 10331, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10331
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10331

Alignment Scores:
Pred. No.: 2,62e-06 Length: 1034
Score: 144.00 Matches: 135
Percent Similarity: 31.73% Conservative: 89
Best Local Similarity: 19.12% Mismatches: 226
Query Match: 4.35% Indels: 256
DB: 10 Gaps: 30

US-09-868-987-1 (1-1864) x US-09-815-242-10331 (1-1034)
Qy 308 TTGGCAATGCTTATTGTTTGTATGAGCGCTATATTATAGATTGGAGGCGTCATGCTTCG 367
|||||
Db 353 LeuValMetTyrLeuPheLeuGlnAsnMetArgAlaThrLeuIleProThrIleAlaVal 372
|||||
Qy 368 GGAGCTGTTCTCTGAATCTTTTGTCTATCTGGCAGCTCTACAGTATTTGGATGCGCCA 427
|||||
Db 373 ProValValLeuLeuGlyThrPheAlaIleLeuAlaPheGlyTyrSerIleAsnThr 392
|||||
Qy 428 CTCACCTTGTGAGGCTCGCTGGGATGTTCTTCTGCTATGGGATGGCGGTAGATGCAAAAT 487
|||||

```

```

Db 393 LeuThrMetPheGlyMet-----ValLeuAlaIleGlyLeuLeuValAspAspAla 409
Qy 488 GTTCTGTTATTCGAAAGAAATCCGAGAGAAATTTTATTGTCTCAAGTCTTT-----AAA 541
|||||
Db 410 IleValValGluAsnVal---GluArgValMetMetGluAspLysLeuProProls 428
|||||
Qy 542 AAATCTGTAGAAAAGATATATCAAGCGCTTTTGAGCCATTTTGTGATTCTTAACCTGACT 601
|||||
Db 429 GluAlaThrGluLysSerMetSerGlnIleGlnGlyAlaLeuVal-----GlyIleAla 446
|||||
Qy 602 ACAGTATTGGCTCAGCACCTCTT-----TTCTTCTTAGATACAGGCGCTATT 649
|||||
Db 447 MetValLeuSerAlaValPheIleProMetAlaPheGlyGlySerThrGlyAlaIle 466
|||||
Qy 650 ---AAAGGGTTTGTCTTTCACATGATTGTTTGGAAATTTTCTTCAATGTTTACGGCTCTT 706
|||||
Db 467 TyrArgGlnPheSerIleThrIleValSerAlaMetAlaLeuSerValLeuValAlaLeu 486
|||||
Qy 707 TTCATGACT----- 715
|||||
Db 487 IleLeuThrProAlaLeuCysAlaThrLeuLysProValSerAlaGluHisGlu 506
|||||
Qy 716 ---AAATTTTCTTCATGCTGTGATGAATAAGACC---CAACATACACACTTGCATATG 769
|||||
Db 507 AsnLysGlyGlyPhePheGlyTyrPheAsnThrThrPheAspHisSerValAsnHisTyr 526
|||||
Qy 770 ATGAAT-----AAAGTTCGTGGGATAAAGCATGTTTCTTG----- 805
|||||
Db 527 ThrAsnSerValGlyLysIleLeuGlySerThrGlyArgTyrLeuLeuIleTyrAlaLeu 546
|||||
Qy 806 -----AGAGATGCAAAAACATTTGGCGCTGTTTCTCGAAGCTGTTTCTTT----- 850
|||||
Db 547 IleValAlaGlyMetValValPheLeuLeuArgLeuProSerSerPheLeuProGluGlu 566
|||||
Qy 851 -----TTAGGTTGGCTGCTCTCGGTTTGGAGCTGGAAT----- 886
|||||
Db 567 AspGlnGlyValPheLeuThrMetIleGlnLeuProAlaGlyAlaThrGlnGluArgThr 586
|||||
Qy 886 ----- 886
Db 587 GlnLysValLeuAspGlnValThrAspTyrTyrLeuLysAsnGluLysAlaAsnValGlu 606
|||||
Qy 887 -----TCCGTTTGGGAATGGATTTTAAAGGA----- 913
|||||
Db 607 SerValPheThrValAsnGlyPheSerPheSerGlyGlnAlaGlnAsnAlaGlyMetAla 626
|||||
Qy 913 ----- 913
Db 627 PheValSerLeuLysProTyrGluArgAsnGlyAspGluAsnSerAlaGluAlaVal 646
|||||
Qy 914 -----GGGTATGCTTTTACCTTTTAAAT 934
|||||
Db 647 IleHisArgAlaLysMetGluLeuGlyLysIleArgAspGlyPheValIleProPheAsn 666
|||||
Qy 934 ----- 934
Db 667 MetProAlaIleValGluLeuGlyThrAlaThrGlyPheAspPheGluLeuIleAspGln 686
|||||
Qy 934 ----- 934
Db 687 AlaGlyLeuGlyHisAspAlaLeuThrGlnAlaArgAsnGlnLeuLeuGlyMetAlaAla 706
|||||
Qy 935 -----CCAAAAGATCGCATCGCATCGCATGTTGCTCAAAATG 970
|||||
Db 707 GlnHisProAlaSerLeuValSerValArgProAsnGlyLeuGluAspThrAlaGlnPhe 726
|||||
Qy 971 CGTGGCAAGTT---GTGCATAAACTACAGGAAGCTGGTCTTCTTCTTAGAGACTTCGGT 1027
|||||
Db 727 LysLeuGluValAspGlnGluLysAlaGlnAlaLeuGlyValSerLeuSerAspIleAsn 746
|||||
Qy 1028 ATTCAACATTTGATCTTCA----- 1048
|||||
Db 747 ---GlnThrIleSerThrAlaLeuGlyGlyThrTyrValAsnAspPheIleAspArgGly 765
|||||

```







```
QY 569 GCTTTTGGAGCCATTTTGGATTCTTAAC----- 595
Db 425 PheArgAlaAlaIleIleGlyThrArgGluIleAlaValProValIleSerMetThrIle 444
QY 596 ---TTGACTACAGTATTGGCTCAGCAGCTCTTTTCTCCTAGATACAGGGCCCTATT--- 649
Db 445 AlaIleAlaValTyrSerProMetAlaLeuMetGlyGlyIleThrGlyThrLeuPhe 464
QY 650 AAAGGGTTGCTTTCACATGATTGAGAAATTTCTCTCAATGTTTACGGCTCTTTTC 709
Db 465 LysGluPheAlaLeuThrLeuAlaGlyAlaValPheIleSerGlyValValAlaLeuThr 484
QY 710 ATGACTAAATTTTCTTCATG-----CTGTGGATG--- 739
Db 485 LeuSerProMetMetSerSerLysLeuLeuLysSerAsnAlaLysProThrTyrMetGlu 504
QY 740 AATAAGACCCCAACATACA-----CAGTTG 763
Db 505 GluArgValGluHisThrLeuGlyLysValAsnArgValTyrGluTyrMetLeuAspLeu 524
QY 764 CATATGATGAATAAGTTTCGTGGGATAAAGCATGATTTCTTGAGAGGATGCAAAAACTT 823
Db 525 ValMetLeuAsnArg-----LysSerMet 532
QY 824 TGGGCTGTTCTGGAAGTGTTTTCTTTTA----- 853
Db 533 LeuAlaPheAlaValIlePheSerThrLeuProPheLeuPheAsnSerLeuSerSer 552
QY 854 -----GGTTGCGTTCCTCGGTTTGGAGCTCGGNAT 886
Db 553 GluLeuThrProAsnGluAspLysGlyAlaPheIleAlaIleGly---AsnAlaProSer 571
QY 887 TCCGTT----- 892
Db 572 SerValAsnValAspTyrIleGlnAsnAlaMetGlnProTyrMetLysAsnValMetGlu 591
QY 893 -----TTGGGAATGATTTTAAAGAGGGTATGCTTTACCTTTAAATCCA 937
Db 592 ThrProGluValSerPheGlyMetSerIleAlaGly---AlaProThrSerAsnSer 609
QY 938 AAGAGCATGCATCAGC-----GATGTTGCTCAATCGCTGGCAAGTT 982
Db 610 SerLeuAsnIleIleThrLeuLysAspTyrLysGluArgSerArgLysGlnSerAlaIle 629
QY 983 GTGCATAAATACAGGAGCTGGTCTTCTCTAGAGACTTCCGTAATCAACATTT--- 1039
Db 630 MetAsnGluIleAsnGluLysAlaLysSerIleProGluValSerValSerAlaPheAsn 649
QY 1039 ----- 1039
Db 650 IleProGluIleAspThrGlyGluGlnGlyProProValSerIleValLeuLysThrAla 669
QY 1040 -----GGATCTTCAGAAAAGATC----- 1057
Db 670 GlnAspTyrLysSerLeuAlaAsnThrAlaGluLysPheLeuSerAlaMetLysAlaSer 689
QY 1058 ---AAAATCTATTTTAGTGAATAAGCTTTAAGCTAT---ACTAAGCAGATACGAGCCCT 1111
Db 690 GlyLysPheIleTyrThrAsnLeuAspLeuThrTyrAspThrAlaGlnMetThrIleSer 709
QY 1112 CTCCTAAAA-----TTAAGCATCATG----- 1132
Db 710 ValAspLysGluLysAlaGlyThrTyrGlyIleThrMetGlnIleSerAsnThrLeu 729
QY 1133 -----AGTGGCGT 1141
Db 730 GlySerPheLeuSerGlyAlaThrValThrArgValAspValAspGlyArgAlaTyrLys 749
QY 1142 TATTGTGGATTGTGTGAGAAAC-----AGCCCTAGATTT 1177
Db 750 ValIleSerGlnValLysArgAspArgLeuSerProGluSerPheGlnAsnTyrTyr 769
```

```
QY 1178 CTCTACGAAACTCTAAACGAAACGCAAAATTTTGGTCAAAAGGTAAACGACCAACTATCG 1237
Db 770 LeuThrAlaSerAsnGlyGlnSerValProLeuSerSerValIleSerMetLysLeuGlu 789
QY 1238 AAGAAA----- 1243
Db 790 ThrGlnProThrSerLeuProArgPheSerGlnLeuAsnSerAlaGluIleSerAlaVal 809
QY 1244 -----ATGCGTTATCAGGCGACC--- 1261
Db 810 ProMetProGlyIleSerSerGlyAspAlaIleAlaTrpLeuGlnGlnAlaThrAsp 829
QY 1261 ----- 1261
Db 830 AsnLeuProGlnGlyTyrThrPheAspPheLysSerGluAlaArgGlnLeuValGlnGlu 849
QY 1262 -----ATCGGGCTTTTAGGAGCTTTTGGCAATCATCTTGCTCTATGTGAGTTTGGCG 1312
Db 850 GlyAsnAlaLeuAlaValThrPheAlaLeuAlaValIleIleIlePheLeuValLeuAla 869
QY 1313 TTGGAATGGCAATATGCTTTTCAGTCCGTATGCGCTTTAATTCATGACCTTTTGGCTACC 1372
Db 870 IleGlnPheGluSerIleArgAspProMetValIleMetIleSerValProLeuAlaVal 889
QY 1373 TGTGCACTCTTG-----TTTATAGCACATTTCTTTTGAAGAAAATTCAAATA 1420
Db 890 SerGlyAlaLeuValSerLeuAsnIleLeuSerPhePheSerIleAlaGlyThrThrLeu 909
QY 1421 GATTTCGAACCCATGCTGCTTTAATGACATGATGTTGGGTATTCATTAAACAATACTTTG 1480
Db 910 AsnIleTyrSerGlnValGlyLeuIleThrLeuValGlyLeuIleThrLysHisGlyIle 929
QY 1481 ATCATTTTTCATGCTATTCGTGAAGATCGCAACGCAACCTGTTTACCCTCATCATGTT 1540
Db 930 LeuMetCysGluValAlaLysGluGluGlnLeuAsnHisGlyLysThrArgIleGluAla 949
QY 1541 TTAGTTAATGATGCCCTTCAAAAGACGCTTCAGCGCAGGTAATGACACAGCTACAACT 1600
Db 950 IleThrHis---AlaAlaLysValArgLeuArgProIleLeuMetThrThrAlaAlaMet 968
QY 1601 CTATCAGTTTTGTTAAATGCTTTGTTTATAGGGCGCTCC-----TCTGCTCTTAAT 1651
Db 969 ValAlaGlyLeuIleProLeuLeuTyrAlaThrGlyAlaGlyAlaValSerArgPheSer 988
QY 1652 TTTGCATTTATATGACCATGAGGATTTCTTAGGAACCTTTATCGTCTCTTTATATTGCA 1711
Db 989 IleGlyIleValIleValAlaGlyLeuSerIleGlyThrIlePheThrLeuPheValLeu 1008
QY 1712 CCA-----CCTGTTGTTGTTTATGTCGCT 1738
Db 1009 ProValValTyrSerTyrValAlaThrGluHisLysProLeuProValPhe----- 1025
QY 1739 AAGAAAATCGCTCA 1753
Db 1026 AspGluAsnLysThr 1030
RESULT 12
US-09-912-020-282
; Sequence 282, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DV1
; CURRENT APPLICATION NUMBER: US/09/912,020
```

```

; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 282
; LENGTH: 1040
; TYPE: PRT
; ORGANISM: E. coli
US-09-912-020-282

Alignment Scores:
Pred. No.: 8.62e-05 Length: 1040
Score: 129.00 Matches: 102
Percent Similarity: 38.45% Conservative: 81
Best Local Similarity: 21.43% Mismatches: 180
Query Match: 3.69% Indels: 114
DB: 10 Gaps: 18

US-09-868-987-1 (1-1864) x US-09-912-020-282 (1-1040)
QY 308 TTGGCANTGCTTATTGATGAGCGTATTAATTAAGATT-----CGAGGC 355
DB 349 MetAlaIleAlaValMetIleIleThreupheLeuArgAsnIleProAlaThr 368
QY 356 GTCAATCGCTCGGAGCTGTTCTTCTGAATCTTTGCTTATCGGAGCTCTACAGTAT 415
DB 369 IleIleProGlyValAlaValProLeuSerLeuIleGlyThrPheAlaValMetValPhe 388
QY 416 TTGATGCGGCACCTACCTTGTCAAGACTCGCTGGGATGTTCTTGCATAGGAGTGGCC 475
DB 389 LeuAspPheSerIleAsnAsnLeuThrIleMetAlaLeuThrIleAlaThrGlyPheVal 408
QY 476 GTATATGCAATGTTCTTGTATTCGAAAGATC-----CGAGAGGAATTT 520
DB 409 ValAspAspAlaIleValIleGluAsnIleSerAlaGlyIleGluValGlyGluLys 428
QY 521 TTATGTCTCAAGTCTTAAATAATCTGTAGAAAAGATATACCAAGCTTTTGGAGCC 580
DB 429 ProLeuAlaIleAlaIleLeuLysGlyAlaGlyGluIleGlyPheThrIleIle----- 445
QY 581 ATTTTGATTTAACTTACTACAGATATGGCTTCAGACATTTCTTTCTTCTTGATACA 640
DB 446 -----SerLeuThrPheSerLeuIleAlaValIleProLeuLeuPheMet 461
QY 641 GGGCCTATTAAAGG-----TTTGCTTGACATGATTGTAATTTTC 685
DB 462 GlyAspIleValGlyArgLeuPheArgGluPheAlaIleThrLeuAlaValAlaIleLeu 481
QY 686 TCTTCATGTTTACGGCTCTTTTCACTAAGATTTTCTTCATGCTGTGANTGAATAAG 745
DB 482 IleSerAlaValAlaSerLeuThrLeuThrPro-----MetMetCysAlaArgMet 498
QY 746 ACCCAACATACACAGTTTCATATGATGATTAATTCGTTGGGATTAACATGATTTCTTG 805
DB 499 LeuSerGlnGluSerLeuArgLysGlnAsnArgPheSerArgAlaSerGluLysMetPhe 518
QY 806 -----AGAGATGCAAAAACCT-----TGGGCT 829
DB 519 AspArgIleIleAlaAlaIleGlyArgGlyLeuAlaLysValIleuAsnHisProThrLe 538
QY 830 GTTTCGAACTGTTTCTTTTAGTTCGCTGCTCGGATTTGAGCTCGAATTC 889
DB 538 uThrLeuSerAla-----AlaLeuSerThrLeuLeuLeuSerValLe 552
QY 890 GTTTTGGAAATGATTTTAAAGAGGGTATGCTTTACT-----TTAATTCAAA 940
DB 552 uLeuThrValPheIleProLysGlyPhePheProValGlnAspAnglyIleIleGlnGln 572
QY 941 GAGCATGCAATCAGCAGATGTTGCTCAATGCTGGCAAAAGTTTGCAATTAACATACAGAA 1000
DB 941 ----- 1000

```

```

DB 572 YThrLeuGlnAlaPro----- 577
QY 1001 GCTGCTCTTTCTTACAGACTTCGCTGATTCAAAATTTGGATCTTCAGAAAAGATCAA 1060
DB 578 -----GlnSerSerSerPheAlaAsnMetAlaGlnArgGlnArgGlnVal 592
QY 1061 ATCATTTTATGATTAAGCTTAAGCTTAATACATACAGACAGATAGAGCCCTCTCCATAAA 1120
DB 592 AlaAspValIleLeuGlnAspProAlaValGlnSer----- 604
QY 1121 TTAACGATCATGACTGCGCTTATGTTGAGATTTGTTCAAGAACAGCCCTAGATTTCTC 1180
DB 605 -----LeuThrSerPheValGlyValAspGlyThr 614
QY 1181 TACGAAAATCTTAACGA-----AACGAAAATTTTGTCAAAAGTTAG 1224
DB 614 rAsnProSerLeuAsnSerAlaArgLeuGlnIleAsnLeuLysProLeuAspGluArgAs 634
QY 1225 CAGCAAACTATCGAAGAAATG-----CGTTATCGAGCGACATCGGCTTTAGAGCTTT 1281
DB 634 pAspArgValGlnLysValIleAlaArgLeuGlnThrAlaValAspLysValProGlyVal 654
QY 1282 GGCATATCATCTTG-----CTCTATGTGAGTTGGCGCTT 1314
DB 654 lAspLeuPheLeuGlnProThrGlnAspLeuThrIleAspThrGlnValSer---ArgThr 673
QY 1315 TGAATGCAATATGCTTTCATGAGCCGATGCGCTTAATTCATGACCTTTGGCTACCTG 1374
DB 673 rGlnIlyrGlnPheThrLeuGlnAlaThrSerLeu-----AspAlaLeuSerThr-- 689
QY 1375 TGCAGCTGCTTTATAGACATTTCTTTTGAAGAAAATCAATAGATTTGCAAGCCAT 1434
DB 690 -----TrpValProGlnLeuMetGluLysLeuGlnGlnLe 701
QY 1435 TGGTCTTATATGACTGATATGCGGTATTCATTAACAAATCTTGATCTTGATGTCG 1494
DB 701 uProGlnLeuSerAspValSerSerAspTrpGlnAspLysGlyLeuValAlaIleValAs 721
QY 1495 TATTCGTGAATGCGCCAGACGACCTGTTTACCCCTATGATGTTTACTGATATGATGC 1554
DB 721 nValAspArgAspSerAlaSerArgLeuGlyIleSerMet---AlaAspValAspAsnAl 740
QY 1555 CTTTCAAAAACGTTCAAGCCGACGGAATGACACAGCTACACT 1600
DB 740 aLeuTrpAsnAlaPheGlyGlnArgLeuIleSerThrIleTrpThr 755

RESULT 13
US-09-815-242-10203
; Sequence 10203, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselebeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

```

```

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: PatSeq for Windows Version 4.0
; SEQ ID NO 10203
; LENGTH: 1040
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10203

Alignment Scores:
Pred. No.:      8,82e-05          Length:      1040
Score:         129.00           Matches:     102
Percent Similarity: 38.45%       Conservative:   81
Best Local Similarity: 21.43%    Mismatches:   180
Query Match:     3.89%          Indels:      114
DB:              10            Gaps:        18

US-09-868-987-1 (1-1864) x US-09-815-242-10203 (1-1040)

QY 308 TTGCGCAATGCTTATTCTTTTGATGCGGTATATTATAGATT-----GGAGGC 355
Db 349 MetAlaIleAlaLeuValMetIleIleTyLeuPheLeuArgAsnIleProAlaThr 368
QY 356 GTCATCGCTTCGGGAGCTGTCTTCGAATCCTTTGCTTATCTGCGGACGCTCACAGTAT 415
Db 369 IleIleProGlyValAlaValProLeuSerLeuIleGlyThyPheAlaValMetValPhe 388
QY 416 TTGGATGCCGCACCTACCTCTGTTCAGACTCGCTGGGATTTCTTGTCTATGGGATGCC 475
Db 389 LeuASPheSerIleAsnAsnLeuThrLeuMetAlaLeuThrIleAlaThrGlyPheVal 408
QY 476 GTAGATGCAAATGTTCTTGTATTCGAAAGAATC-----CGAGAGGAATTT 520
Db 409 ValASPAlaIleValValIleGluAsnIleSerArgTyIleGluLysGlyLulys 428
QY 521 TTATTGTCTCAAAGTCTTAATAAATCTGTAGAAAAAGGATATACCAAGCCTTTTGGAGCC 580
Db 429 ProLeuAlaAlaLeuLysGlyAlaGlyGluIleGlyPheThrIleIle----- 445
QY 581 ATTTTTGATTTCTAACTTGACTACAGTATTGGCTCAGCACCTCTTTCTTCTCCTAGATACA 640
Db 446 -----SerLeuThrPheSerLeuIleAlaValLeuIleProLeuLeuPheMet 461
QY 641 GGSCCTATTAAAGG-----TTTGCTTTGACATTCATTATAGGAATTTTC 685
Db 462 GlyASPileValGlyArgLeuPheArgGluPheAlaIleThrLeuAlaValAlaIleLeu 481
QY 686 TCTTCAATGTTTACGCGCTCTTTTCATGACTAAATTTTCTTCATCTCTGGATGAATAAG 745
Db 482 ILeSerAlaValValSerLeuThrLeuThrPro-----MeuCysAlaArgMet 498
QY 746 ACCCAACATACACAGTTGTCATATGATGAATAAGTTCTGGGGATAAACGATGTTCTTGT 805
Db 499 LeuSerGlnGlusLeuArgLysGlnAsnArgPheSerArgAlaSerGluLysMetPhe 518
QY 806 -----AQAGATGCAGAAAACCT-----TGGGCT 829
Db 519 AspArgIleIleAlaIleAlaTyrgLyArgGlyLeuAlaLysValLeuAsnHisProTrp-Le 538
QY 830 GTTCTTCGGAAGTCTTTTCTTTTAGGTTCCGTTCTCGGGTTTGGAGCCTGGAAATTC 889
Db 538 uThrLeuSerVal-----AlaLeuSerThrLeuLeuLeuSerValLe 552
QY 890 GTTTTGGGAATGGAATTTTAAAGAGGGTATGCCTTTTACCCT-----TTAATCCAAA 940
Db 552 uLeuTrpValPheIleProLysGlyPhePheProValGlnAspAsnGlyIleIleGlnGI 572
QY 941 GAGCATGGCATCAGCGATGTTCCTNAATGCTCGCAAGATTTGTGCATAAATACTACAGAA 1000

```

```

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-C2-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11516
; LENGTH: 1028
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11516

Alignment Scores:
Pred. No.: 0.000111 Length: 1028
Score: 128.00 Matches: 73
Percent Similarity: 43.10% Conservative: 77
Best Local Similarity: 20.58% Mismatches: 142
Query Match: 3.86% Indels: 56
DB: 10 Gaps: 17

US-09-868-987-1 (1-1864) x US-09-815-242-11516 (1-1028)
QY 131 TTGAAATATCATGCGATGCTCGAGGAATTTACCCAGCGTGAAGTACGAACTGCC 190
DB 296 LeuYsh1st1eglnAl1leSerProSerTyrlu1leatgProphleuAspThr 315
QY 191 TCAGATTTAAATCTGAGCGATGCTTTTGTCCGAGTTCTCAGTGAAGACATC 250
DB 316 SerTyrl1ealgnr-----Ser1legluAspVal 325
QY 251 TCTTGTGATCTTGGAATAAACAATGACACAGGCAATATCTCAGATGCTGTGGCTG 310
DB 326 LysPheAspLeu-----lleuug1yAla1le-----Leu 335
QY 311 GCATGCTTATGTTTTCATGATGAGGTATATATAGATTGAGGCGTC-----ATCGCT 364
DB 336 Ala1leVal1leuVal1Phe-----AlaPheLeuArgSer1yThr1leThLeuValSer 353
QY 365 TCGGAGCGTGTCTTCGATCTTTCGATCTTTCGAGCGAGCTCTACAGTATTTGGATCG 424
DB 354 Ala1leSer1lePro1leSer1leMetGlyThrPheAla1leu1leGlnTrpMetGlyPhe 373
QY 425 CCAGCTCCTTGTCAGACGTCGCTGGAGTGTCTTCTATAGGGATGCGGTAGATCA 484
DB 374 SerLeuAsnMetLeuThMetValAla1leuThrLeuAla1leGly1le1le1leAsp 393
QY 485 AATGTTCTTATTCGAGAGATCCGAGAGATTTTATGTCCTCAAGTCTTAAAAA 544
DB 394 Ala1leVal1leGlnAsn1leHis1ySylsLeuGluMetSer1yAsArg1yS 413
QY 545 TCTGTAGAAAAAGATATACCAAGCTTTTGGAGCATTTTGTATCTTCACTTGACTACA 604
DB 414 AlaSer1yrg1ugly---Val1ySglu1leGlyPheAla-----LeuValAla 428
QY 605 GTATTGCGCTCAGACCTCTTTCTTCTCATATACAGGCGCTATTAAGG----- 655
DB 429 l1eSerAlaMetLeuLeuSerValPheValPro1leGlyAsnMet1ySgl1y1le1leGly 448
QY 656 -----TTTGCTTTCGATTTAGGATTTTTCGATTTTCTCTCAATGTTTACG 700
DB 449 ArgPhePheGlnSerPheGly1leThrValAla1leuAla1le1leuSer1yValVal 468
QY 701 GCTCTTTCATGACTAAATTTTCTTCATGCTGTGATGAGATTAAGCCCAATACACAG 760
DB 469 ValVal1Thr1le1lePheMetValSerSerValValAsn---ProArg1SerArg 487
QY 761 TTGCATATGCAATGAAAC-----TTGCGGGGATTAACATGATTTCTTGAGAGATCC 814
DB 488 PheTyMetTrpSerGluProPhePhe1ySalal1eGln1ySerArg1yThr1ySleu1eu 507
QY 815 AAAAACTTTGGGCTGTTCTCGAAGTGTTTTCTTTTATGTTGGGCTGCTCGGGTTT 874

```

```

DB 508 Gln-----TrpVal1leuAsn1ySylsLeu1le1lePhe1leAlaVal1leuValPhe 525
QY 875 -----GGAGCGCTG-----AATTCGCTTTGGGATGATTTT-----AAA 910
DB 526 Val1ySgluPheValAla1leSer1ySleuGlyMetGluPheMetLeu1ySgluAspArg 545
QY 911 GAGGATATGCTTTTACCTTTAATCCAAAAGACATGCGATCAGC---GATGTGCTCA 967
DB 546 GlyArgPheLeuValTrpLeu1ySal1ySPro---GlyValSer1leAspTyMetThr 564
QY 968 ATGCGTGGCAAGTT-----GTGCATTAACCTACAGAACTGCTTCTTCTTCT 1015
DB 565 Gln1ySer1ySer1lePheGln1ySalal1eGlu1yS1eAspGluValGlnPheThr 584
QY 1016 AGAGACTTCGATTTCAACATTTGATCTTCGAAAAGATCAAAATCTATTATGAT 1075
DB 585 LeuGlnVal1yGlyTy1yGlyThrThrglnAsnProPhe1ySalal1y1lePheValGlnLeu 604
QY 1076 AAGCTTTAAGCTATTAAGCAG 1099
DB 605 LysProLeu1ySgluArg1ySyls 612

RESULT 15
US-09-912-020-283
; Sequence 283, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Twack, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DVI
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 283
; LENGTH: 1025
; TYPE: PRT
; ORGANISM: E. Coli
US-09-912-020-283

Alignment Scores:
Pred. No.: 0.000318 Length: 1025
Score: 123.50 Matches: 130
Percent Similarity: 33.33% Conservative: 105
Best Local Similarity: 18.44% Mismatches: 243
Query Match: 3.73% Indels: 227
DB: 10 Gaps: 28

US-09-868-987-1 (1-1864) x US-09-912-020-283 (1-1025)
QY 245 ACGATCTCTCTGATCTTGGAATAAACAATGACACAGCATTTATCTCAGATGCTGT 304
DB 326 Thr1leArgAlaSerLeuGluGluValGlnGlnThrLeu---lle1leSer----- 341
QY 305 GCGTGGCAATGCTTATGTTTTCATGAGCGTATATATATGATTGGA-----GGCGTC 358
DB 342 ---ValAlaLeuVal1leLeuValValPheLeuPheLeuArgSerGlyArgAlaThr1le 360
QY 359 ATGCGCTCGGAGCTGTTCTTCTGATCTTTCGATCTTTCGATCTTCAGAGCTCTACATATTG 418
DB 361 l1eProAlaValSerValProValSer1leu1leGlyThrPheAla1leMetTy1rLeuCyS 380

```

QY 419 GATCGCCACCTACCTTGTCTCAGGACTCGCTGGGATGTCTTCTATATGGGATGGCCGCTA 478  
 Db GlyPheSerLeuAsnAsnLeuSerLeuMetAlaLeuThrIleAlaThrGlyPheValVal 400  
 QY 479 GATGCAAAATGTTCTTATTCGAAAGATTCGAGAGGAATTT-----TTA 523  
 Db AspAspAlaIleValValLeuGluAsnIleAlaArgHisLeuGluAlaGlyMetLysPro 420  
 QY 524 TTGCTCTCAAGCTCTTAAATAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGCCATT 583  
 Db LeuGlnAlaLeuGlnGlyThrArgGluValGlyPheThr-----Val 435  
 QY 584 TTGATTTCACTTACCTACAGTATGGCTCAGCAGCTCTTTTCTTCTAGATACAGGG 643  
 Db LeuSerMetSerLeuSerLeuValAlaValPheLeuProLeuLeuMetGlyGlyLeu 455  
 QY 644 CCT-----ATTAAAGGGTTTCTTTGTGACATTGATTTAGGAATTTCTCTTCAATG 694  
 Db ProGlyArgLeuLeuArgGluPheAlaValThrLeuSerValAlaIleGlyIleSerLeu 475  
 QY 695 TTTACGGCTCTTTTCATGACTAAATTTTCTTCTCATGCTGTGATGAATAAGACC----- 748  
 Db LeuValSerLeuThrLeuPro---MetMetCysGlyTrpMetLeuLysAlaSerLys 494  
 QY 749 -----CAACATACACAGTTGATGATGATGAATAAG-----TTCGTGGGATAAAGCATGAT 799  
 Db ProArgGluGlnLysArgLeuArgGlyPheGlyArgMetLeuValAlaLeuGlnGlnGly 514  
 QY 800 TTTCTGAGAGGATGCAGAAAATCTTTGGGCTGTTCTCGAAGT-----GTTTTT 847  
 Db TyrGlyLysSerLeuLys-----TrpValLeuAsnHisThrArgLeuValGlyValVal 532  
 QY 848 CTTTATGAGTTGGCTGCTCTCGGTTTGGAGCTCGAATTCGGTT----- 892  
 Db LeuLeuGlyThrIleAlaLeuAsnIleTrpLeuTyrlleSerIleProLysThrPhePhe 552  
 QY 893 -----TTGGGA----- 898  
 Db ProGluGlnAspThrGlyValLeuMetGlyGlyIleGlnAlaAspGlnSerIleSerPhe 572  
 QY 899 -----ATGGATTTT----- 907  
 Db GlnAlaMetArgGlyLysLeuGlnAspPheMetLysIleIleArgAspProAlaVal 592  
 QY 908 -----AAAGGAGGATGCTCCCTTTACC 928  
 Db AspAsnValThrGlyPheThrGlySerArgValAsnSerGlyMetMetPheIleThr 612  
 QY 929 TTTAATCAAAGACATGGCATCAGCGATGTGTCAAAATCGTGGCAAAAGTTGTGCAT 988  
 Db LeuLysProArgAspGluArgSerGluThrAlaGlnGlnIleIleAspArgLeuArgVal 632  
 QY 989 AAACATA---CAGGAAGCTGGT-----CTTCTCTAGACACTTCGCTATT 1030  
 Db LysLeuAlaLysGluProGlyAlaAsnLeuPheLeuMetAlaValGlnAspIleArgVal 652  
 QY 1031 -----CAACATTTGGATCTTCAGAAAGATCAAAATCTATTTTAGTGATAAAGCTTTA 1084  
 Db GlyGlyArgGlnSerAsnAlaSerTyrlnTyrlnLeuLeuSerAspAspLeuAlaAla 672  
 QY 1085 -----AGCTATACAGCATACGACCTCTCTCTAAATTAACGATCATG----- 1132  
 Db LeuArgGluTrpGluProLysIleArgLysLysLeuAlaThrLeuProGluLeuAlaAsp 692  
 QY 1133 -----AGC 1135  
 Db ValAsnSerAspGlnAspAsnGlyAlaGluMetAsnLeuValTyrlnAspArgAspThr 712  
 QY 1136 TGGCGTTATTGGGATTTGTTCTCAGAACAGCGCTAGATTCTCTACGGAACCTCTAAA 1195  
 Db MetAlaArgLeuGlyIleAspValGln-----AlaAlaAsnSerLeu 726

QY 1196 CGAACCGCAAAATTTTGGTCAAGGTAAGCAGCAAACTATCGAAGAAAATG----- 1246  
 Db LeuAsnAsnAlaPheGlyGlnArgGlnIleSerThrIleTyrlnProMetAsnGlnTyrln 746  
 QY 1247 -----CGTTATCAGGCGACCATCGGCTTTTAGGAGCTTTG 1282  
 Db LysValValMetGluValAspProArgTyrlnThrGlnAspIleSerAlaLeuGluLysMet 766  
 QY 1283 GCAATCATC-----TTGCTCTATGTAGTTTGGCTTTTGAATGCAATAT 1327  
 Db PheValIleAsnAsnGluGlyLysAlaIleProLeuSerTyrlnAlaLysTrpGlnPro 786  
 QY 1328 GCT----- 1330  
 Db AlaAsnAlaProLeuSerValAsnHisGlnGlyLeuSerAlaAlaSerThrIleSerPhe 806  
 QY 1330 ----- 1330  
 Db AsnLeuProThrGlyLysSerLeuSerAspAlaAlaIleAspArgAlaMetThr 826  
 QY 1331 -----TTCAGTCCGCTATCGCTTTTAAATCAT 1357  
 Db GlnLeuGlyValProSerThrValArgGlySerPheAlaGlyThrAlaGlnValPheGln 846  
 QY 1358 GACCTTTTGGCTACTGCTGCTGCTGCTTTTATAGCACATCTTCTTTTGAAGAAAATTCAA 1417  
 Db GluThrMetAsnSerGlnValIleLeuIleAlaIleAlaIleAlaThrValTyrlleVal 866  
 QY 1418 ATAGATTG----- 1426  
 Db LeuGlyIleLeuTyrlnGluSerTyrlnHisProLeuThrIleLeuSerThrLeuProSer 886  
 QY 1427 CAAGCCATTGCTGCTTTA----- 1444  
 Db AlaGlyValGlyAlaLeuLeuAlaLeuGluLeuPheAsnAlaProPheSerLeuIleAla 906  
 QY 1445 -----ATGACTGTATTCGGGTATTTCATAACAATACTTTGATCATCTTTTGCAT 1492  
 Db LeuIleGlyIleMetLeuLeuIleGlyIleValLysLysAsnAlaIleMetMetValAsp 926  
 QY 1493 CGTATTCGTGAAGATCGCAAGCGAACCTGTTTACCCCTATGTCATGTTTATAGTTAATGAT 1552  
 Db PheAlaLeuGluAlaGlnArgHisGlyAsnLeuThrProGlnGluAlaIlePheGlnAla 946  
 QY 1553 GCCCTTCAAAGACGTTTCAGCCGCGCACGTAATGACAAACAGCTACAACTCTATCAGTTTGT 1612  
 Db CysLeuLeuArg---PheArgProIleMetMetThrThrLeuAlaAlaLeuPheGlyAla 965  
 QY 1613 TTAATGCTTTTGTATTAGCGGC---TCCTCTCTCTTTTAAATTTTGCATTTATTATGACC 1669  
 Db LeuProLeuValLeuSerGlyAspGlySerGluLeuArgGlnProLeuGlyIleThr 985  
 QY 1670 ATA-----GGGATTTCTTAGGAACCTTTATCGTCTCTTTATATGACCACTCTGTTG 1723  
 Db IleValGlyGlyLeuValMetSerGlnLeuLeuThrLeuTyrlnThrProValValTyrln 1005  
 QY 1724 TTGTTTATGCTCCGT 1738  
 Db LeuPhePheAspArg 1010  
 RESULT 16  
 US-09-815-242-5111  
 ; Sequence 5111, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Karl L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard

```

; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5111
; LENGTH: 1062
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5111

```

## Alignment Scores:

```

Pred. No.: 0.00408 Length: 1062
Score: 122.50 Matches: 65
Percent Similarity: 39.86% Conservative: 51
Best Local Similarity: 22.34% Mismatches: 114
Query Match: 3.70% Indels: 61
Gaps: 11

```

US-09-868-987-1 (1-1864) x US-09-815-242-5111 (1-1062)

```

QY 104 GTCAGACCCCTATTATTAACGTCCTCAATGCAAAATCATGCCAGTGTCTCAGGAAATTT 163
Db 290 ValAlaIleProIlePheGlnArgProGlySerAsnAlaIleGluIleSer---AsnLeu 308
QY 164 ACCACCGGTGAGAGAGAACTGCCCTCAATTTAAATGTGAGGAGATCTTTGTT 223
Db 309 ValArgGluGlyMetAlaGluLeuLysHisSerPheProGlnGly---MetAspTyr-- 326
QY 224 CCCAGGTTCTCAGTGAAGAGACGATCTTGTGATCTTGGGAAAAAACAATGTACACA 283
Db 327 ---SerIleValIlyrAprProThIle-----PheValArg 337
QY 284 GGCATTATCTCAGCA-----TGCTGTGCTTGGCAATGCTTATT 322
Db 338 GlYserIleGluAlaValAlaValHisThrLeuPheGluAlaLeuValIleValIleVal 357
QY 323 GTTTGATGAGCGTATATTATAGATTGGAGCGTCATCGTTGGGAGCGTCTTCTG 382
Db 358 ValIleLeuPheLeuGlnThrTyrPArgAlaSerIleIleProLeuAlaAlaValProAl 377
QY 383 AATCTTTGCTTATCTGAGCAGCTCTACAGATTGATGAGCGCATCACTTGTCAAGA 442
Db 378 SerIleuIleGlyThrPheAlaValMetHisMetLeuGlyPheSerLeuAsnAlaLeuSer 397
QY 443 CTCGCTGGATGTTGTTCTGATAGGGAGTGGCCGTAGTGCAAATGTTCTTGTATTCGA 502
Db 398 LeuPheGlyLeuValIleValIleAlaIleGlyIleValIValaPhePalaIleValIleValGlu 417
QY 503 AGAATCCGAGGGAATTTTATTGTCTCAAAAGTCTTAAATAATCTGTGAAGAAAGGAT 562
Db 418 AsnValGluArgAsnIleGlyLeuGlyLeuLysProValGluAlaThrLysArgAlaMet 437
QY 563 ACCAAGGCTTTGGAGCATTTTGTATCTTAAGTACTGACTACAGATAGTGGCTGACACTT 622
Db 438 ArgGluValThrGlyProIle-----IleAlaThrAlaLeu 449

```

```

QY 623 CTTTC-----TTCCTAGATACAGGCGCTATTAAAGG----- 655
Db 450 ValLeuCyseAlaValPheIleProThrAlaPheIleSerGlyLeuThrGlyGlnPheTyr 469
QY 656 -----TTTGCTTTGACATGATTTTATGGAATTTCTCTCAATGTTTACGGCTCTTTC 709
Db 470 ArgGlnPheAlaLeuThrIleAlaIleSerThrValIleSerAlaPheAsnSerLeuThr 489
QY 710 ATGACT-----AAATTTTCTTC 727
Db 490 LeuSerProAlaLeuAlaAlaValLeuLeuLysGlyHisIleGluProLysAspArgPhe 509
QY 728 ATGCTGGATGATTAAGACCCACATACACAGTTG---CATATGATGATTAAGTTC--- 781
Db 510 SerValPheLeuAspLysLeuLeuGlySerTyrPheLeuPheArgProPheAsnArgPhePhe 529
QY 782 -----GTGGGATTAAGCATGATTTCTTGGAGATGCAAAAA 820
Db 530 AspArgAlaSerHisGlyTyrValGlyThrValAsnArgValLeuArgLysSerSerIle 549
QY 821 CTTGGGCTGTTTCTGGAAGTGTCTTTTCTTTA 853
Db 550 AlaLeuLeuValTyrGlyGlyLeuMetValLeu 560

```

## RESULT 17

```

US-09-895-913A-10
; Sequence 10, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.

```

```

; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in th
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 761
; TYPE: PRF
; ORGANISM: Helicobacter pylori
US-09-895-913A-10

```

## Alignment Scores:

```

Pred. No.: 0.000399 Length: 761
Score: 122.00 Matches: 57
Percent Similarity: 45.52% Conservative: 75
Best Local Similarity: 19.66% Mismatches: 124
Query Match: 3.68% Indels: 34
Gaps: 13

```

US-09-868-987-1 (1-1864) x US-09-895-913A-10 (1-761)

```

QY 305 GGCTTGCCATGCTTATTTGATGAGCGTATATTATGATTGGAGCGTC----- 358
Db 332 GlyAlaIleLeuAlaValLeuValAlaPheAlaPheLeuArgAsnGlyThrIleThrLeu 351
QY 359 ATGCTTCGGAGCTGTTCTTGAATCTTTGCTTATCTGAGCAGCTCAAGTATTTG 418
Db 352 ValSerAlaIleSerIleProIleSerIleMetCylThrPheAlaLeuIleGlnTrpMet 371
QY 419 GATGCGCACTACCTTGTGAGACTGCGTGGATTTGTTGCTATGGGATGCGCGTA 478
Db 372 GlyPheSerLeuAsnMetLeuThrMetValAlaLeuThrLeuAlaIleGlyIleIleIle 391
QY 479 GATGCAATGTCTTGTATTCGAAAGATCCGAGAGAAATTTTATTGTCTCAAGTCTT 538

```

```

Db 392 AspaapAlaIleValIleGluAsnIleHisIysLeuGluMetGlyMetSerIys 411
Qy 539 AAAAAATCTAGAAAAAGGATATACCAAGGCTTTTGGAGCCATTTTTCATTTCTAACTTG 598
Db 412 ArgIysAlaSerTyrGluGlyValArgGluIleGlyPheAla-----Leu 426
Qy 599 ACTACAGATTGGCTCAGCACTCTTTCTCTCTAGATACAGGCCCTATTAAAGGG--- 655
Db 427 ValAlaIleSerAlaMetLeuLeuSerValPheValProIleGlyAsnMetIysGlyIle 446
Qy 656 -----TTTCTTTGACATTGATTTTAGGAATTTTCTCTCAATG 694
Db 447 IleGlyArgPhePheGlnSerPheGlyIleThrValAlaLeuAlaIleAlaLeuSerTyr 466
Qy 695 TTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTGTGTGATGAATPAAGACCAACAT 754
Db 467 ValValValValThrIleIleProMetValSerSerValValValAsn---ProArgHis 485
Qy 755 ACACAGTTGCATATCATGAATAAG-----TTCGTGGGGATAAAGCATGATTTCTTGAGA 808
Db 486 SerArgPheTyrValTrpSerGluProPhePheIysAlaLeuGluSerArgTyrThrLys 505
Qy 809 GGATCAAAAAAATTTTGGCTGTCTTCGAGTGTCTTTTCTTTAGGTGGCTGCTCTC 868
Db 506 LeuLeuGln-----TrpValLeuAsnHisIysIleIleIleSerIleAlaValValLeu 523
Qy 869 GGGTTT---GGAGCTCG---AATTCGGTTTTTGGGAATGATTTT----- 907
Db 524 ValPheValGlySerLeuPheValAlaSerIysIleGlyMetGluPheMetLeuLysGlu 543
Qy 908 ---AAAGAGGATATCGCTTTACCTTTTATCCAAAGAGCATGGCATCAGC---GATGTT 961
Db 544 AsparGlyIa-gPheLeuValTrpLeuLysAlaIysPro---GlyValSerIleAspTyr 562
Qy 962 GCTCAATGCTGGCGCAAGTGTGTCATAAACTACAGGAAGCTGGTCTTCTCTAGAGAC 1021
Db 563 MetThrGlnLysSerIysIlePheGlnLysAlaIleGluLysHisAlaGluValGluPhe 582
Qy 1022 TTCCGTATTCAACA---TTTGGATCTTCAGAA-----AAGATCAAAATCTATTTT 1069
Db 583 ThrThrLeuGlnValGlyTyrGlyThrGlnAsnProPheLysAlaLysIlePheVal 602
Qy 1070 AGTGATAAGCTTTAAGCTATCTACTAAGCAG 1099
Db 603 GlnLeuLysProLeuLysGluArgIysLys 612

```

## RESULT 18

```

US-09-815-242-5194
; Sequence 5194, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

```

```

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5194
; LENGTH: 1043
; TYPE: PR1
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5194

Alignment Scores:
Pred. No.: 0.000456 Length: 1043
Score: 122.00 Matches: 154
Percent Similarity: 33.83% Conservative: 117
Best Local Similarity: 19.23% Mismatches: 260
Query Match: 3.68% Indels: 270
DB: 10 Gaps: 37

US-09-868-987-1 (1-1864) x US-09-815-242-5194 (1-1043)
Qy 74 CGTATGCTGTAGTGTGACGGTTATATGTCAGCAGCCCTATTTTAAAGCTCCCATTTG 133
Db 269 ArgLeuGluLeuGlyLysGluSerTyrAsnIleSerArgLeuAsnGlyThrPro--- 287
Qy 134 AAAAATCATGCGATGCTCAGGGGAAATTTACCCACCGTGAAGTGAGCAAACTCGCTCA 193
Db 288 -----ThrValGlyAlaIle-----GlnLeuSerProGlyAlaAsn 300
Qy 194 GATTTAAATCTGGAGCGATG-----TCTTTTGTTCCTCC 226
Db 301 AlaIleGlnThrAlaThrLeuValLysGlnArgLeuAlaGluLeuSerAlaPhePro 320
Qy 227 GAG-----GTTCTCAGTGAAGAGACGATCTCTCTGATCTTGGGAAAAA 271
Db 321 GluAspMetGlnTyrSerValProTyrAspThrSerArgPheValAspValAlaIleGlu 340
Qy 272 CAATGTACACAAAGCATTATCTCAGCA-----TGCTGTGCTTGGCAATGCTTATTGTT 325
Db 341 LysValIleHisThrLeuIleGluAlaMetValLeuValPheLeuValMetPheLeuPhe 360
Qy 326 TTGATGACGCTATATATAGATTTTGGAGGCGTCATCGCTTCGGGAGCGTGTCTTCTGAAT 385
Db 361 LeuGlnAsnValArgTyrThrLeuIleProSerIleValValProValCysLeuLeuGly 380
Qy 386 CTTTGTCTTATCTGGGCGCTCTACAGTATTGATGGCCACTCACCTTGTCTAGGACTC 445
Db 381 ThrLeuMetVal-----MetTyrLeuLeuGlyPheSerValAsnMetMetThrMet 397
Qy 446 GCTGGGATTGTTCTGTATGGGATGGCGCTAGATGCAAAATGTTCTTGTATTTCGAAGA 505
Db 398 PheGlyMetValLeuAlaIleGlyIleLeuValAspAlaIleValValValGluAsn 417
Qy 506 ATCCGAGAGGAATTTTATTCTCTCAAGTCTTAAA-----AAATCTGTAGAAAAAGGA 559
Db 418 Val---GluArgIleMetAlaGluGluGlyIleSerProAlaGluAlaThrValYsAla 436
Qy 560 TATACCAAGGCTTTTGGAGCCATTTTGTATCTTAACCTTACAGTATTGGCTCAGCA 619
Db 437 MetLysGlnValSerGlyAlaIleVal-----GlyIleThrLeuValLeuSerAlaVal 454
Qy 620 CTT-----CTTTCTTCCTA-----GATACAGGGGCTATT---AAAGGGTTTGTCTTG 664
Db 455 PheLeuProLeuAlaPheMetAlaGlySerValGlyValIleTyrGlnPheSerVal 474
Qy 665 ACATTTGATTTTAGGAATTTTCTCTTCAATGTTTACGGCTCTTTTCATGACTAAA----- 718
Db 475 SerLeuAlaValSerIleLeuPheSerGlyPheLeuAlaLeuThrPheThrProAlaLeu 494
Qy 719 -----TTTTTTC----- 724

```







; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5330  
; LENGTH: 801  
; TYPE: PR1  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-5330

Alignment Scores:  
Pred. No.: 0.00104 Length: 801  
Score: 118.00 Matches: 43  
Percent Similarity: 48.30% Conservative: 42  
Best Local Similarity: 24.43% Mismatches: 73  
Query Match: 3.56% Indels: 18  
DB: 10 Gaps: 7

US-09-868-987-1 (1-1864) x US-09-815-242-5330 (1-801)

QY 1253 CAGGCGACCATCGGCTTTAGGAGCTTTGGCAATCATCTTGTCTATGTAGTGGC 1312  
Db 357 LysAlaSerLeuGlyThrIleValAlaIleValIleLeuLeuPhe-----LeuArg 374  
QY 1313 TTTGAATGGCAATATGCTTTCAGTCCGCTATCGCTTTAAATCATGACCTTTTGGCTACC 1372  
Db 375 AsnIleArgThrThrAlaIleSerIleIleSerIleProLeuSerLeuMetAlaLeu 394  
QY 1373 TGTGAGCTTGTGTTTATAGCACATTTCTTTTGAAGAAAAATTCAAATAGATTTGCAAGCC 1432  
Db 395 IleAlaLeuLys-----LeuSerAspValSerLeuAsnIleLeuThr 408  
QY 1433 ATTGGTGTCTTAATGACGTATGTTGGGTATTCATTAACAATCTTTGATCATTTTGTAT 1492  
Db 409 LeuGlyAlaLeuThrValAlaIleGlyArgValIleAspAspSerIleValValValGlu 428  
QY 1493 CGTATTCGTGAA-----GATCGCAAGCGAACCTGTTTACCCCTATGCATGTTTAA 1543  
Db 429 AsnIleTyrArgArgLeuThrAspSerGluGluGlnLeu-----LysGlyGluAsnLeu 446  
QY 1544 GTTAATGATGCTTCAAAAGACGTTTACGCGCACGGTAATGACACACAGCTACAACTCTA 1603  
Db 447 IleIleSerAlaThrThrGluValPheLysProIleMetSerThrLeuValThrIle 466  
QY 1604 TCAGTTTGTGTTAATGCTTTTATATAGCGGCTCC-----TCTGCTTTAAAT---TTT 1654  
Db 467 IleValPheLeuProLeuValPheValSerGlySerValGlyGluMetPheArgProIle 486  
QY 1655 GCATTTATTATGACCATAGGATCTTCTAGGAACCTTTATCGTCTCTTATATATTCACCA 1714  
Db 487 AlaLeuAlaIleAlaPheSerLeuLeuAlaSerLeuLeuValSerIleThrLeuValPro 506  
QY 1715 CCTCTGTTGTTTATGTCGCTAAA-----GAAAAATCGCTCAAAA 1756  
Db 507 AlaLeuAlaAlaThrLeuPheLysLysGlyValLysArgArgAsnLys 522

## RESULT 20

US-09-815-242-10072  
; Sequence 10072, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815.242  
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,948  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10072  
; LENGTH: 1049  
; TYPE: PR1  
; ORGANISM: Escherichia coli  
US-09-815-242-10072  
Alignment Scores:  
Pred. No.: 0.00131 Length: 1049  
Score: 117.50 Matches: 138  
Percent Similarity: 30.95% Conservative: 109  
Best Local Similarity: 17.29% Mismatches: 278  
Query Match: 3.55% Indels: 273  
DB: 10 Gaps: 33  
US-09-868-987-1 (1-1864) x US-09-815-242-10072 (1-1049)  
QY 44 AATGACACATATTTCTGCAACCGTGGATGCGTATGCTGTAGTGTGACGGTTATATG 103  
Db 282 AsnGlyGlnProAlaSerGlyLeuGlyIleLysLeuAlaThrGlyAlaAsnAlaLeuAsp 301  
QY 104 GTCAGACGCTTATTTAAACGTCCTCCATTCGAAATCATGCCAGTGTCTCAGGAAATTT 163  
Db 302 ThrAlaAlaIle----- 306  
QY 164 ACCCAGCGTGAAGTGAGCAAACTCGCCTCAGATTTAAATCTGGAGCGATGCTCTTTGTT 223  
Db 307 ---ArgAlaGluLeuAlaLysMetGluProPheProPheGlyLeuLysIleValTyr 325  
QY 224 CCC-----GAGGTTCTCAGTGAAGAGACGATCTCTCTCGATCTT 262  
Db 326 ProTyrAspThrThrProPheValLysIleSerIleHisGluValVal----- 341  
QY 263 GGGAAAAACATGTACACAGGCAATTCACAGCATGCTGTGCGCTTGGCAATGCTTATT 322  
Db 342 -----LysThrLeuValGluAlaIleIleLeuValPhe-----LeuValMetTyrLeu 357  
QY 323 GTTTTGTATGACGCTATATTATAGATTTGGAGGCGTCATCGCTTCGGGAGCTGTTCTCTG 382  
Db 358 PheLeuGlnAsnPheArgAlaThrLeuIleProThrIleAlaValProValValLeuLeu 377  
QY 383 AATCTTTTGTCTTATCTGGGAGCTCTACAGATTTTGGATGCGCCACTCACCTTGTCTCAGGA 442  
Db 378 GlyThrPheAlaValLeuAlaAlaPheGlyPheSerIleAsnThrLeuThrMetPheGly 397  
QY 443 CTCGCTGGGATGTTGTTTGTCTATGGGATGCGGCTAGATGCAAAATGTTCTTGTATTGAA 502  
Db 398 Met-----ValLeuAlaIleGlyLeuLeuValAspAlaIleValValValGlu 414  
QY 503 AGATCCGAGGGAATTTTATGCTCTCAAGTCTT-----AAAAATCTGTAGAAAA 556  
Db 415 AsnVal---GluArgValMetAlaGluGlyLeuProLysGluAlaThrArgLys 433  
QY 557 GGAATATACCAAGGCTTTTGGAGCCATT-----TTTGATTTCACTTGACCTACAGTATTG 610  
Db 434 SerMetGlyGlnIleGlnGlyAlaLeuValGlyIleAlaMetValLeuSerAlaValPhe 453  
QY 611 GCCTCAGACCTTCTTTTCTTCTCTAGATACAGGCGCTATT---AAAGGTTTGTCTTGACA 667



; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 5186  
 ; LENGTH: 1018  
 ; TYPE: PRF  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-815-242-5186

Alignment Scores:  
 Pred. No.: 0.00164 Length: 1018  
 Score: 116.50 Matches: 136  
 Percent Similarity: 34.15% Conservative: 101  
 Best Local Similarity: 19.60% Mismatches: 219  
 Query Match: 3.52% Indels: 239  
 DB: 10 Gaps: 32

US-09-868-987-1 (1-1864) x US-09-815-242-5186 (1-1018)

QY	305	GGCTTGGCAATGCTTATGTTTGGATGAGCGTATATATTAGATTGGAGGC-----	355
DB	338	GlyGluAlaValLeuIleValIleValValValPheLeuPheLeuGlyAlaPheArgSer	357
QY	356	---GTCATCGCTTCGGGAGCTCTTCTTCTGAATCTTTTCTTATCTGGCGAGCTCTACAG	412
DB	358	ValLeuIleProValValThrIleProLeuSerMetIleGlyValLeuPheMetGln	377
QY	413	TATTGGATCGCCACTCACCCTGTGAGGACTCGCTGGGATGTTCTTCTGATGGGATG	472
DB	378	AlaMetGlyTyrSerIleAsnLeuThrLeuLeuAlaMetValLeuAlaIleGlyLeu	397
QY	473	GCCGTAGATGCAATGTCTGTATTCCAAAGATCCGA-----GAGGAA-----	517
DB	398	ValValAspAlaIleValValValGluAsnIleHisArgHisIleGluGluGlyLys	417
QY	517	-----	517
DB	418	ProProPheGluAlaAlaLeuGluGlyAlaArgGluIleAlaValProValValSerMet	437
QY	518	-----TTTTTA-----TTGTCTCAAAAGT	535
DB	438	ThrIleThrLeuAlaAlaValTyrAlaProIleGlyPheLeuThrGlyLeuThrGlyAla	457
QY	536	CTTAAAAAATCTGAGAAAAGGATATACCAAGGCTTTTGGAGCCATTTTCAATCTAAC	595
DB	458	LeuPheLys-----GluPheAlaPheThrLeuAla--GlyAlaValIleIleSerGly	474
QY	596	TTGACTACAGTATTGGCCTCAGCACTT-----	622
DB	475	IleValAlaLeuThrLeuSerProMetMetCysSerArgLeuLeuArgHisGluGluAsn	494
QY	623	-----CTTTCTCTTC-----	631
DB	495	ProSerGlyLeuAlaHisArgLeuAspLeuIlePheGluGlyLeuLysGlnArgTyrGln	514
QY	632	-----CTAGATACAGGCTATATAAGGGTTTGGCTTTGTCATGTGAT	673
DB	515	ArgAlaLeuHisGlyThrLeuAspThrArgProValValLeuValPheAlaValLeuVal	534
QY	674	TTAGGAATTTTCTCT-----TCAATGTTTACGGCTCTTTTCATGACT-----	715
DB	535	LeuAlaLeuIleProValLeuLeuMetPheThrLysLysGluLeuAlaProGluGluAsp	554
QY	716	AAATTTTCTTCATGCTGTGGATGAATAGACCCCAACATACACAGTTGCATATGATGAT	775
DB	555	GlnGlyIleValPheLeuMetThrAsnSerProGlnThrAlaAsnLeuAspTyrLeuAsn	574
QY	776	AAG-----TTCGTGGGATAAGCATGATTTCTTGAGAGGATGCCAAAACCTT	823



```

Db 363 AsnIleSerValHisPheValLeuIleProThrLysAsnLeuTrpCysPheAlaThr 382
Qy 1505 GATCCGCAAGCAACCTTTTACCCTATGCTGTTTGTAGTAAATGATGCTCTTCAAAAG 1564
Db 383 ThrAlaSerTyrLeuLeuLeuValLeuHisTyrPheVal-----AlaLysLys 399
Qy 1565 ACGTTCACCCGCGGTAATGACACACCTACAACTCTATCAGTTTG-----TTAATGCTT 1621
Db 400 LysTyrAlaTyrAspGluValAlaIleSerThrPheValLysValLeuAlaLeuValVal 419
Qy 1622 TTGTTTATAGCGGCTCTCTGCTCTT-----AATTTTGCATTATTATGACCATAGG 1675
Db 420 ValTyrThrGlyLeuMetThrValPheValGlySerIleTyrPheValTrpSerLeuGly 439
Qy 1676 ATTCTTCTAGGAACCTTATGCTCTTATATGTCACCACTCTGTTGTTGTTATG--- 1732
Db 440 IleAlaValLeuValValTyrAlaIleTyrPheArgLysGluLeuThrValAlaLeuAsn 459
Qy 1733 GTCCGTAAAGAAATCGCTCAAAA 1756
Db 460 ThrPheArgGluLysArgSerLys 467

RESULT 23
US-10-139-876-20
; Sequence 20, Application US/10139876
; Patent No. US20020123481A1
; GENERAL INFORMATION:
; APPLICANT: Oliviero, Salvatore
; TITLE OF INVENTION: C-Fos Induced Growth Factor (Figf) And Dna Encoding Same
; FILE REFERENCE: 35784/205172
; CURRENT APPLICATION NUMBER: US/10/139,876
; CURRENT FILING DATE: 2002-05-07
; PRIOR FILING DATE: 1998-03-18
; PRIOR FILING DATE: 1996-09-30
; PRIOR FILING DATE: 1996-06-13
; PRIOR FILING DATE: 1995-09-29
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 597
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-139-876-20

Alignment Scores:
Pred. No.: 0.00474 Length: 597
Score: 111.00 Matches: 57
Percent Similarity: 34.51% Conservative: 31
Best Local Similarity: 22.35% Mismatches: 88
Query Match: 3.47% Indels: 79
DB: 12 Gaps: 12

US-09-868-987-1 (1-1864) x US-10-139-876-20 (1-597)
Qy 798 TCATGCTTT---ATCCCAACCACTTATCATCATGCACTGTGTATGTTGGTCTTA 742
Db 290 ThrCysTyrGlyIleAlaThrAsnValAsnValPheCysArgArgLysIleHisLeuLeu 309
Qy 741 TTCATCCACGACGATCAAGAAAATTTAGTCATGAAACA----- 703
Db 310 -----GluInLysThrThrLeuIleSerArgAsnGlnLeuSerValGlyHisThr 326
Qy 702 -----GCCCTAAACATGTAAGAAAATTTCTTAAATCAATGTC 664
Db 327 CysLeuThrLysIleValAlaSerValSerValLysHisValProLysIle----- 344
Qy 663 AAGCAACCCCTTTAATAGGCCCTGTATCTAGGAAGAAAAGAGTGTGAGGCCCAATACT 604

```

```

Db 345 -----SerSerThrProLysThr 350
Qy 603 GTAGTCAAGTTAGTAATCAAAAATGGCT-----CCAAAAGCCTTTGGTATATCTCTTTTCT 550
Db 351 AlaValAlaLeuSerAlaLysLysValTrpArgProAlaAla----- 364
Qy 549 ACAGATTTTAAAGACTTTTGAGACAATAAAATTCCTCGGATTCCTTCCGAATACAAGA 490
Db 365 -----ArgSerThrSer 368
Qy 489 ACATTTCATCTACGGCCATCCCATAGCAAGAACAAATCCCGAGAGTCTCTGACAAAGTG 430
Db 369 TyrPheThrGlnThrProAlaAlaValArgThrAspAlaProPheIleProAspHisVal 388
Qy 429 AGTGGCCGATCAAAATCTGTAGAGTCCCGAGATAGCAAAAGATTCCAGAAAGACAGCT 370
Db 389 GlnValAlaLysGlnHisValGlnSerIleAlaAlaPheGlnArg---ArgLysGlyLeu 407
Qy 369 CCCGAAGGATGAGCGCTCCAAATCTATATATATAGCTCATCAAAACAAATAGCAATGCC 310
Db 408 ProArgGlyProThrAlaGluArgIle-----LeuAspSerAlaPheGlnValPro 424
Qy 309 AAGCCA-----CAGCATGCTGAGATAATGCTTGTGTACATTGTTTTCCTCCAAAGA 259
Db 425 HisProCysHisPheGlnHisAlaLeuProSerCysCysHisCysPhePheProArg 444
Qy 258 TCAGAAGAGATCGTCTCTTCACTGAGAACCTCGGGAACAAAGACATCGCTCCAGATTTT 199
Db 445 Cys-LysLysAsnProPheTyrThrAlaProGlnIleGln---ThrAsnLeuProPheTh 463
Qy 198 AATCTGAGGCGAGTTTGCTCCTCCTCAGGTGGGTAAATTTCCCTGAGACACTGGCATGA 139
Db 463 rProAlaLysGlu-----SerLeuValHisTrpMetSe 474
Qy 138 TTTTTCATGGGCGGTTTAAATAGGCTGCTGACCATATTAAC 96
Db 474 rSerSerCysArgCysLeu-----CysAlaProArgAsn 485

RESULT 24
US-09-815-242-10968
; Sequence 10968, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10968

```

```

; LENGTH: 396
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-10968

```

Alignment Scores:	
Pred. No.:	0.0637
Score:	109.00
Percent Similarity:	33.43%
Best Local Similarity:	18.44%
Query Match:	3.2%
DB:	10
	14
	Gaps:
	14
	Mismatches:
	97
	Conservative:
	52
	Matches:
	366
	Length:
	366

QY	1037	TTTGAGCTTCAGAAAGATCAAAATC---TATTTAGGATTAAGAGCTTTAAGCTATCT	10933
Db	25	PheksnThlrThrgluPheValProValAMetLeuSerAspIleAlaGlnSerPhe	44
QY	1094	AAGCAGATACGAGCGCTCTCTCTAAATTAACGATCATGAGCTGGCGTTATTTGGAGAT	11533
Db	45	MetGlnThrlaAspThrgluLeuMetMetThrValTyrAlaTrpThr-----Val	61
QY	1154	GTTGTCAGAAACAGCGCTACATTTCTC---TACGAAACTTAAACA-----	11988
Db	62	LeuIleMetSerLeuProAlaMetLeuAlaThrGlyAsnMetGluThrGlySerLeu	81
QY	1198	-----	11988
Db	82	IleLysLeuPheIleIlePheIleValGlyHisIleLeuSerValIleAlaTrpAsn	101
QY	1199	-----AAGCAAAATTTTGGTCAAG-----	12198
Db	102	TrpIleLeuLeuAlaArgMetCysIleAlaLeuAlaHisSerValPheTrpSerIle	121
QY	1220	GTAAGCAGC-----AACTATGCAAGAAATGCGTTACAGCGCAGCAGCTCGGCTT	12707
Db	122	ThrIaSerLeuValMetArgIleSerProLysIleLysIleThrGlnAlaLeuGlyMet	141
QY	1271	TTAGGA-----GCTTGGCAGATCATCTTG-----CTCATGTAGT	13068
Db	142	LeuAlaIleGlyThrIlaLeuAlaIleThrIleLeuGlyLeuProIleGlyArgIleValGly	161
QY	1307	TTGGCGTTTGAATGCAATATGCTTTACAGCCGCTATGCGCTTAAATTCAGACCTTTG	13668
Db	162	GlnLeuValGlyTrpArgValThrPheGlyIleIleAlaValLeuAlaLeuSerIleMet	181
QY	1366	-----	13668
Db	182	PheLeuIleIleArgLeuLeuProAsnLeuProSerLysAsnAlaGlySerIleAlaSer	201
QY	1367	-----GCTACCTGTGCACTCTTG-----	13847
Db	202	LeuProLeuLeuAlaLysArgProLeuLeuLeuTrpLeuTyrValThrThrAlaIleVal	221
QY	1385	TTTATAGCACATTC-----TTTTGAAGAAATTCAAATAGATTTTGCAAGCC	14322
Db	222	IleSerAlaHisPheMetThrAlaTyrThrTyrIleGluProPheMetIleAspValGlyHis	241
QY	1433	ATTGCT-----GCTTAAATACCTGTATGGGGATTCATTAAACAATCT	14777
Db	242	LeuAspProAsnPheAlaThrAlaValLeuLeuValPheGlyPheSerGlyIleAlaAla	261
QY	1478	TTGATCATTTTGAATCGATATCGGAAGATCGCAAGCAACTGTTTACCCCTTAGCAT	15337
Db	262	SerLeuLeuPheAsnArgLeuTyrArg-----PheAlaProThrLys	275
QY	1538	GTTTATAGTTATATGATGCCCTTCAAAAGACGTTACGCGCAGCGTAATGCAACAGCTCA	15977
Db	276	PheIleValIleSerMetSerLeuLeuMetPheSerIleLeuLeuLeuPheSerThr	295
QY	1598	---ACTCATATAGTTTGTGTAATGCTTTTGTGTTATAGCGGC-----	16368

Db 74 ValPheLysGlyProSerArgSerLysLeuArgThrValIleIleCysAsnLeuAlaPhe 93  
Qy 449 GGGATTGTTCTGCTATGGGATGCCCGTAGATGCAATGTTCTTGATTCGAAGAATC 508  
Db 94 AlaIleCysMetMetLeuGlyThrIleuThrAlaLysThrProLeuValPhe----- 110  
Qy 509 CGAGAGGAATTTTATTGCTCTCAAGTCTTAAATAATCTGAGAAAAAGGATATACCAAG 568  
Db 111 -----GlyMetThrLeu 114  
Qy 569 GCTTTTGAGCCATTTTGAATCTTAACCTGACTACAGTATTGGCCTCAGCA----- 619  
Db 115 LeuIleValThrValIleProPheThrIlePheThrAlaLeuLysIleAlaGlyProSer 134  
Qy 620 CTTCTTTTCTCTAGATAGAGG-----CTATT----- 649  
Db 135 SerThrPheIleValThrPheSerLeuProIleAsnLeuProIleAlaProGluLeu 154  
Qy 650 -----AAGGGTTGCTTGTACATGATTTTAGAATTTCTCTCTTAATGTTTACG 700  
Db 155 AlaLeuTyArgGlyPheAlaIle---LeuValGlyGlyIleLeuAlaThrMet----- 171  
Qy 701 GCTCTTTTCATGACTAATTTTCTTCATGCTGTGATGAATAGACCAACATACACAG 760  
Db 172 -----MetValLeuIleThrIleValPheSerLysAsnLysAlaGluGlnAla 188  
Qy 761 TTGCAT-----ATGATGAATAAGTTGCGGGGATAAAGCATGAT----- 799  
Db 189 IleGlnAsnAspPheLysLeuIleSerLysLeuLeuHisThrTyAsnAspLysSerAla 208  
Qy 800 TTCTTGAGAGGATGCAAA-----AAACTTTGGGCT 829  
Db 209 PheLeuLysValAlaLysThrAlaValAspSerPheLysAlaSerAspLysLeuLeu 228  
Qy 830 GTTCTGGAAGTGTTCCTTTAGTTGCGTGTCTCGGGTTGGAGCCGGAATCC 889  
Db 229 ThrSerThrSerSer-----AsnAsp 235  
Qy 890 GTTCTTGGGAATGATTTTAAAGAGGGTATGCTTTTACCTTTAATCCAAAGAGCATGCC 949  
Db 236 LysLeuSerArgArgPheGln-----LysLeuLeuLeuLeuHisThrSerAlaGlnGly 253  
Qy 950 ATC---ACGATGTTGCTCAATGCGTGGCAAAAGTTGTCATAACTACAGGAAGCTGGT 1006  
Db 254 IleTySerGluLeuGluLeuAsnAlaLysGlnIleArgProLeuProAspGluLeu 273  
Qy 1007 CTTTCTCTAGAGCTTCGGTATTCACAACTTTGGATCTTCAGAAAAGATCAAAATCTAT 1066  
Db 274 IleGluMetMetAspHisIleIleAlaGlnLeuAspAsnSerAspGlu---AsnValArg 292  
Qy 1067 TTTAGTGATAAAGCTTTAAGCTATATAAGCAGATACGA-----GCCTCTCTCCTA 1117  
Db 293 TyrTrpArgLysGluValThrValThrGluGluPheGlnAsnLeuPheAsnHisIleLeu 312  
Qy 1118 AAATTAACGATCATGAGCTGGCGTTATTGTGGGATTTGTGAGAAACAGCGCTAGATTT 1177  
Db 313 LysIleAspGluMet-----ValHisAlaAsnGluAlaArgIle 325  
Qy 1178 CTCTACGGAATCTTAACGAAACCCAAATTTTGGTCAAGGTAAGCAGCAAACTATCG 1237  
Db 326 AlaTyGluAlaAspMetArgLysProLeuTySerLysArgIleTyGlnAsnLeuThr 345  
Qy 1238 -----AAGAAATGCGTTATCAGGCGCATCGGGCTTTTAGA 1276  
Db 346 LeuAspSerIleValPheArgAsnThrLeuArgIleThr----- 358  
Qy 1277 GCTTTGGCAATCATCTGCTATGATGAGTTTGGCTTTGAATGCAATATGCTTTTACGT 1336  
Db 359 AlaIleMetMetIleAlaIlePheIleAlaLeuMetPheAsnPheGluLysAlaTyTrp 378  
Qy 1337 GCCGTATGCGCTTTAATTCATGACCTTTTGGCTACTGCTGCGAGTCTTGTATTATACCAT 1396

Db 379 IleProLeuSerAlaHisThrIleLeuLeuGlyThrSerThrIleHisAlaIleGluArg 398  
Qy 1397 TTCTTTTGAAGAAATTCAAATAGATTTTCAAGCCATTGGTGTCTTAATGACTGTATTG 1456  
Db 399 GlyMetAlaArg-----GlyLeuGlyThrIleLeu 408  
Qy 1457 GGGTATTTCATTAAACAATACCTTTCATCATTTTCAT-----CGTATTTCGT 1501  
Db 409 GlyValLeuValLeuSerValIleLeuLeuPheSerIleProThrProValAlaValIle 428  
Qy 1502 GAAGATCGCCCAAGGAACCTGTTTACCCCTATGATGTTTGTAGTTTAAATGATGCTTCAA 1561  
Db 429 LeuMetGlyIleAlaAlaLeuPheThrGluAlaLeuValGlyAlaAsnTyAlaIleAla 448  
Qy 1562 AAGACGTTTACGCCCGCACGGTAATGACACACAGCTACAACTCTATCATGTTTGTAAATGCTT 1621  
Db 449 ValValPhe-----IleThrIleGlnValIleLeuMetAsn 460  
Qy 1622 TTGTTTATAGCGGCTCTCTGCTCTTTAATTTTGCATTT----- 1660  
Db 461 GlyLeuAlaSerGlnAsnLeuThrIleAsnIleAlaPheProArgValIleAspValAla 480  
Qy 1661 -----ATTATGACCATAGGATTT---CTTCTAGGAACCTTTATCGTCTCTCTTAT 1705  
Db 481 MetGlyIleValIleAlaIleIleGlyLeuPheValLeuGlyGlnArgThrAlaSerAla 500  
Qy 1706 ATGCACACACCTCTGTTGTTGTTTATGTCGCTAAAGAA 1744  
Db 501 LeuLeuProAsnValMetAlaGluValValArgLysGlu 513

RESULT 26  
US-09-815-242-12682  
; Sequence 12682, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12682  
; LENGTH: 651  
; TYPE: PRN  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12682

Alignment Scores: 0.00993 Length: 651  
Pred. No.: 108.00 Matches: 110  
Score:

Percent Similarity: 35.62% Conservative: 87  
 Beest Local Similarity: 19.69% Mismatches: 202  
 Query Match: 3.26% Indels: 154  
 Gaps: 26

US-09-868-987-1 (1-1864) x US-09-815-242-12682 (1-651)

```

QY 305 GGCCTGGCAATGCTTATGTTTGATGAGCGTATAT-----TATAGATTGGA 352
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 42 GlyLeuLeuMetIleIleProAlaIleIleGlyTyrLeuCysGlyAsnProGlnPheGly 61
QY 353 GGGGTACGCTTCGAGGAGCTGTCTCTGATCTTTGCTTACTGAGAGCTACAG 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 62 LeuLeuValAlaThrGlyThrLeu-----AlaHisIleTyr 73
QY 413 TATTTGGATGCGCCA-----CTCACCTTGTGAGAGCTGCT-- 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 74 ValPheLeuGlyProSerArgSerLysLeuArgThrValIleIleCysAsnLeuAlaPhe 93
QY 449 GGGATTGCTTGTGATGAGGAGGCGGTAGATGCAAAATGTTCTGTATTCGAAAGATC 508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 94 AlaIleCysMetLeuGlyThrLeuThrAlaLysThrProLeuValPhe----- 110
QY 509 CGAGAGAAATTTTATGTCTCAAAAGCTTAAATACTGAGAAAAGGATATACCAG 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111 -----GlyMetThrLeu 114
QY 569 GCTTTTGAGACCATTTTGAATCTTACTGACATACAGATATGCGCTCAGCA----- 619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 115 LeuIleValThrValIleProPheTyrIlePheThrAlaLeuLysIleAlaGlyProSer 134
QY 620 CTCTTTTCTTCCCTAGATACAGG-----CCTATT----- 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 135 SerThrPhePheIleValIThrPheSerLeuProIleAsnLeuProIleAlaProGlu 154
QY 650 -----AAAGGTTTGCTTGATTTAGAAATTTTGAATTTCTCTTCATGTTTACG 700
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 155 AlaLeuTyrArgGlyPheAlaIle-----LeuValGlyGlyIleLeuAlaThrMet----- 171
QY 701 GCTCTTTTCATGACTAAATTTTCTTCATGCTGTGATGATGAAGAACCAATACACAG 760
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 172 -----MetValLeuIleThrIleValPheSerLysAsnLysAlaGluGlnAla 188
QY 761 TTGCAT-----ATGATGAATTAAGTTCGTGGGATTAACATGAT----- 799
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 189 IleGlnAsnAspPheLysLeuIleSerLysLeuLeuHisThrTyrAsnAspLysSerAla 208
QY 800 TTCTTGAGAGATGCAAA-----AAACTTTGGGCT 829
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 209 PheLeuLysValAlaLysThrAlaValAspSerPheLysAlaSerAspLysLeuIle 228
QY 830 GTTTCGTGAAGTGTCTTTCTTTAGGTTCGCTGCTCGGGTTTGAGCCTGGAATTC 889
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 229 ThrSerThrSerSer-----AsnAsp 235
QY 890 GTTTTGGGAATGATTTTAAAGAGGCTATGCCCTTTACCTTTAATCCAAAAGACATGCC 949
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 236 LysLeuSerArgArgPheGln-----LysLeuLeuLeuLeuHisThrSerAlaGlnGly 253
QY 950 ATC---AGCATGTTGCTAAATGCGGCAAGTTGCGTAAACTGAAACAGAGCTGT 1006
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 254 IleTyrSerGluLeuLeuGlnAlaLysGlnIleArgProLeuProAspGluLeu 273
QY 1007 CTTCCTCTAGAGACTTCGATTCAAACATTTGATCTTCAGAAAGATCAAAATCTAT 1066
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 274 IleGluMetMetAspHisIleIleAlaGlnLeuAspAsnSerArgGlu---AsnValArg 292
QY 1067 TTTAGTATAAAGCTTTAAAGCTATATACAGATACGA-----GCCTCTCTCCTTA 1117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 293 TyrTrpArgLysGluValThrValThrGlnGluPheGlnAsnLeuPheAsnHisIleLeu 312
QY 1118 AAATTAACGATCATGACCTGCGCTTATTTGGAGTTGTGTCAAAACAGGCTAGATT 1177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 313 LysIleAspGluMet-----ValHisAlaAsnGlnAlaArgIle 325
QY 1178 CTCTACGAAACTCTTAAAGCAACGCAAAATTTTGTCTCAAGTAAAGCAGCAACTATAG 1237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 326 AlaTyrGluAlaAspMetArgLysProLeuTyrSerLysArgIleTyrGlnAsnLeuThr 345
QY 1238 -----AAGAAATGCGCTTATCAGGCGACCATCGGGCTTTAGGA 1276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 346 LeuAspSerIleValAlaPheArgAsnThrLeuArgTyrThr----- 358
QY 1277 GCTTTTGCAATCATCTTGTCTATGATGAGTTTTCGCTTGAATGCAATATGCTTCAGT 1336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 359 AlaIleMetMetIleAlaIlePheIleAlaLeuMetPheAsnProGlnLysAlaTyrTrp 378
QY 1337 GCGGTATGCGCTTAAATTCATGACCTTTTGGCTACCTGTGCACTCTTGTATGACACAT 1396
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 379 IleProLeuSerAlaHisThrIleLeuLeuGlyThrSerThrIleHisAlaIleGluArg 398
QY 1397 TTCTTTTGAGAAATTCAAATATAGATTTCAGACCATTTGCGCTTATATGACTGATAG 1456
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 399 GlyMetAlaArg-----GlyLeuGlyThrIleLeu 408
QY 1457 GGGTATTCATTTAAACATATCTTGTATCATTTTGTAT-----CGTATTCGT 1501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 409 GlyValLeuValLeuSerValIleLeuLeuPheSerIleProThrProValAlaValIle 428
QY 1502 GAAGATCGCCAGCGAAGCTGTTTACCCCTATGACATGTTTATGATATGATGCCCTTCAA 1561
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 429 LeuMetGlyIleAlaIleAlaLeuPheThrGluAlaLeuValGlyAlaAsnTyrAlaIleAla 448
QY 1562 AAGACGTCAGCGCGACCGGTAATGACACAGTACACTCTATCAGCTTTGTTATGCTT 1621
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 449 ValValPhe-----IleThrIleGlnValIleLeuMetAsn 460
QY 1622 TTCTTTATAGCGCGCTCTCTGTCTTAAATTTTGCATT----- 1660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 461 GlyLeuAlaSerGlnAsnLeuThrIleAsnIleAlaPheProArgValIleAspValAla 480
QY 1661 -----ATTATGACCATAGGATTT---CTTCAGGAACCTTATGCTCTCTTTAT 1705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 481 MetGlyIleValIleAlaIleIleGlyLeuPheValLeuGlyGlnArgThrAlaSerAla 500
QY 1706 ATTGACCAACCTGCTGTGTTGTTATGTCGCCGTAAGAA 1744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 501 LeuLeuProAsnValMetAlaGluValArgLysGlu 513

```

RESULT 27  
 US-09-815-242-13719  
 ; Sequence 13719, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Karl L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; TITLE OF INVENTION: Prokaryotes  
 ; FILE REFERENCE: ELITRA, 011A  
 ; CURRENT APPLICATION NUMBER: US/09/815,242  
 ; PRIOR FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625



; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 13719  
; LENGTH: 1049  
; TYPE: PRT  
; ORGANISM: Salmonella typhi  
US-09-815-242-13719

Alignment Scores:  
Pred. No.: 0.0122 Length: 1049  
Score: 108.00 Matches: 122  
Percent Similarity: 31.62% Conservative: 100  
Best Local Similarity: 17.38% Mismatches: 240  
Query Match: 3.26% Indels: 240  
DB: 1.0 Gaps: 29

US-09-868-987-1 (1-1864) x US-09-815-242-13719 (1-1049)

```
QY 308 TTGCAATGCTTATGTTTGTGATGAGCGTATATATAGATTGGAGGCGTCATCGCTTCG 367
D 353 LeuValMetTyrLeuPheLeuGlnAsnPheArgAlaThrLeuIleProThrIleAlaVal 372
QY 368 GGAGCTGTTCTTCTGAATCTTTGCTTATCTGGGAGGCTCTACAGATTATTGGATCGGCA 427
D 373 ProValValLeuLeuGlyThrPheAlaValLeuAlaPheGlyPheSerIleAsnThr 392
QY 428 CTCACCTTGTGAGACTCGCTGGGATGTTCTTGTATGGGATGGCGTAGATGCAAT 487
D 393 LeuThrMetPheGlyMet-----ValLeuAlaIleGlyLeuValAspAla 409
QY 488 GTTCTTGTATTCGAAGAATCCGAGAGAAATTTTATGTCTCAAGTCTT-----AAA 541
D 410 IleValValGluAsnVal---GluArgValMetThrGluGluGlyLeuProProlys 428
QY 542 AAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGCCATTTTGAATCTAACTTGACT 601
D 429 GluAlaThrArgLysSerMetGlyGlnIleGlnGlyAlaLeuVal-----GlyIleAla 446
QY 602 ACAGTATTGGCTCAGCACTTCTT-----TTCTTCTTAGATACAGGCGCTATT 649
D 447 MetValLeuSerAlaValPheIleProMetAlaPhePheGlyGlySerThrGlyAlaIle 466
QY 650 ---AAAGGTTTGTCTGACATTGATTTAGGAATTTTCTTCAATGTTTACGCTCTT 706
D 467 TyrArgGlnPheSerIleThrIleValSerAlaMetAlaLeuSerValLeuAlaLeu 486
QY 707 TTCATGACT-----715
D 487 IleLeuThrProAlaLeuCysAlaThrMetLeuLysProValAlaLysGlyAspHisGly 506
QY 716 -----AAATTTTCTCATGCTGTGGATGAATAAGACCAACATACAGTGTGCAT--- 766
D 507 GluGlyLysGlyPhePheGlyTrpPheAsnArgLeuPheAspLysSerThrHis 526
QY 766 -----766
D 527 TyrThrAspSerValTrpAsnIleLeuArgSerThrGlyArgTyrLeuLeuLeuTyrLeu 546
QY 767 -----ATGATGAATAAGTTCTGGGGGATAAAGCATGATTTCTTG----- 805
D 547 IleIleValValGlyMetAlaTyrLeuPheValArgLeuProSerSerPheLeuProasp 566
QY 805 -----805
D 567 GluAspGlnGlyValPheLeuThrMetValGlnLeuProAlaGlyAlaThrGlnGluArg 586
QY 806 -----AGAGGATGC 814
```

```
D 587 ThrGlnLysValLeuAspGluValThrAspTyrTyrLeuAsnLysGluLysAlaAsnVal 606
QY 815 AAAAATCTTTGGGCTGTTCTTGAAGTGTCTTTCTT-----TTAGGTTC 859
D 607 GluSerValPheAlaValAsnGlyPheGlyPheAlaGlyArgGlyGlnAsnThrGlyIle 626
QY 860 GTTCTCTCGGTTTGGAGCTGG-----AATTCGTTTGGGA 898
D 627 AlaPheValSerLeuLysAspTrpAlaAspArgProGlyGluLysAsnLysValGluAla 646
QY 899 ATG-----GATTTTAAAGGAGGCTATGCTTTACCTTT 931
D 647 IleThrGlnArgAlaThrAlaAlaPheSerGlnIleLysAspAlaMetValPheAlaPhe 666
QY 932 AAT-----934
D 667 AsnLeuProAlaIleValGluLeuGlyThrAlaThrGlyPheAspPheGluLeuIleAsp 686
QY 934 -----934
D 687 GlnAlaGlyLeuGlyHisGluLysLeuThrGlnAlaArgAsnGlnLeuPheGlyGluVal 706
QY 935 -----CCAAAAGAGCATGCGCATCAGCGATGTTGCTCAA 967
D 707 AlaLysTyrProAspLeuLeuValGlyValArgProAsnGlyLeuGluAspThrProGln 726
QY 968 ATGCTGCGCAAGTT---GTGCATAAACTACAGAAAGCTGCTTCTTCTTAGAGACTTC 1024
D 727 PheLysIleAspIleAspGlnGluLysAlaGlnAlaLeuGlyValSerIleSerAspIle 746
QY 1025 CGTATT-----CAAACTTTGGATCTTCAGAAAGATCAAAATCTATTTAGTGAT 1075
D 747 AsnThrThrLeuGlyAlaAlaTrpGlyGlySer-----TyrValAsnAsp 761
QY 1076 -----AAAGCTTTAAGCTATCTACTAAGCAGATACGAGCCTCTCTC 1114
D 762 PheIleAspArgGlyArgValLysValTyrValMetSerGluAlaLysTyrArgMet 781
QY 1115 CTAATAATTAACGATCATGAGCTGGCTTATTGTGGGATTGTTGTGCAAAACAGGCT--- 1171
D 782 LeuProAspAspIleAsnAspTrpTyrValArgGlySerAspGlyGlnMetValProPhe 801
QY 1172 -----AGATTTCTCTACGAACTCTCAA-----1195
D 802 SerAlaPheSerSerSerArgTrpGluTyrGlySerProArgLeuGluArgTyrAsnGly 821
QY 1196 -----CGAAACGCA 1204
D 822 LeuProSerMetGluIleLeuGlyGlnAlaAlaProGlyLysSerThrGlyGluAlaMet 841
QY 1205 AAATTTTGGTCAAGGTAAAGCAGCAAACTATCGAAG-----AAA 1243
D 842 AlaMetMetGluGluLeuAlaSerLysLeuProSerGlyIleGlyTyrAspTrpThrGly 861
QY 1244 ATGCTTATCAGGCGCACCATCGGG-----CTTTTAGGAGCTTTGGCAATC 1288
D 862 MetSerTyrGlnGluArgLeuSerGlyAsnGlnAlaProAlaLeuTyrAlaIleSerLeu 881
QY 1289 ATCTTGCTCTATGTG-----AGTTTGGCTTTGATGCAATATGCTTTCAGTCC 1339
D 882 IleValValPheLeuCysLeuAlaLeuTyrGluSerTrpSerIleProPheSerVal 901
QY 1340 GTATGCGCTTAAATTCATGACCTTTTGCTACCTGTCAGCTTTGTTTATAGCACATTTC 1399
D 902 MetLeuValVal-----ProLeuGlyValIleGlyAlaLeuLeuAlaAlaThrPhe 918
QY 1400 TTTTGAAGAAAATTCAAATAGATTTCAGCCCATTTGGTCTTTAATGACTGTATTGGGG 1459
D 919 -----ArgGlyLeuThrAsnAspValTyrPheGlnValGlyLeuLeuThrThrIleGly 936
QY 1460 TATTCAATTAACAATACTTTGATCATTTTGTATCGTATTCTGTAA-----GATCGCCAA 1513
D 937 LeuSerAlaLysAsnAlaIleLeuIleValGluPheAlaLysAspLeuMetAspLysGlu 956
```



QY 1625 TTTATAGCGGCTCTCTGCTTTAATTTTCATTTATTATG---ACCATAGGATCTTT 1681  
 |||||  
 Db 433 PheIleIleGly-----PheLeuPheProTyrIleThrGlyAlaIleGlyLeuAla 449  
 |||||  
 QY 1682 CTAGAACTTTATCGTCTCTATATTCACACCTCTGTTGTTGTTTATGTC 1735  
 |||||  
 Db 450 LeuGlyGlyTyrValPheLeuValPheAlaGlyLeuValLeuPheIleLeu 467

## RESULT 29

US-10-095-139-14  
 ; Sequence 14, Application US/10095139  
 ; Patent No. US20020165357A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Curtis, Rory A.J.  
 ; APPLICANT: Silos-Santiago, Immaculada  
 ; APPLICANT: Millennium Pharmaceuticals, Inc.  
 ; TITLE OF INVENTION: 38554, 57301, and 58324, Human Organic  
 ; TITLE OF INVENTION: Ion Transporters and Uses Therefor  
 ; FILE REFERENCE: MF101-017P1RNM  
 ; CURRENT APPLICATION NUMBER: US/10/095,139  
 ; CURRENT FILING DATE: 2002-03-11  
 ; PRIOR APPLICATION NUMBER: 60/275,172  
 ; PRIOR FILING DATE: 2001-03-12  
 ; NUMBER OF SEQ ID NOS: 24  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 14:  
 ; LENGTH: 487  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: consensus  
 ; US-10-095-139-14

## Alignment Scores:

Pred. No.: 0.0319 Length: 487  
 Score: 102.50 Matches: 110  
 Percent Similarity: 35.67% Conservative: 68  
 Best Local Similarity: 22.04% Mismatches: 169  
 Query Match: 3.09% Indels: 152  
 DB: 9 Gaps: 20

US-09-868-987-1 (1-1864) x US-10-095-139-14 (1-487)

QY 596 TTGACTACGATTGGCTCAGCACTCTTTTCTCTAGATACAGGCTATTAAAGGG 655  
 |||||  
 Db 3 LeuValAlaAlaLeuGlyGlyPheLeuPheGlyTyrAspThrGlyValIleGly 22  
 |||||  
 QY 656 TTTCGTTTGCAATTCATTT----- 673  
 |||||  
 Db 23 Phe---LeuAlaLeuIleAspPheLeuPheArgPheGlyLeuLeuThrSerSerGlyAla 41  
 |||||  
 QY 674 -----TTAGGAATTTCTCTCAATGTTTACGGCTCTTTTCATGACTAAATTTTC 724  
 |||||  
 Db 42 LeuAlaGluLeuGly---TyrSerThrValLeuThrGlyLeuValValSerIlePhe 60  
 |||||  
 QY 725 TTCATGCTGTGGATCAATPAGACCCACATACACAGTTGTCATGATCAATAAGTTTCGTG 784  
 |||||  
 Db 61 LeuGlyArgLeuIleGly-----SerLeuPheAla 70  
 |||||  
 QY 785 GGGATAAGCATGATTTCTGAGAGGATGCAAAACCTTTGGCTGTTTCTGGAGTGTT 844  
 |||||  
 Db 71 GlyIysLeuGlyAspArgPhe---GlyArgIysIysSerLeuLeuIleAlaLeuValLeu 89  
 |||||  
 QY 845 TTCTTTTGTGTTGGTCTCGCTCGGTTTGGAGCTGGA-----ATTCCGTTTGG 895  
 |||||  
 Db 90 PheValIleGly-AlaLeuLeuSerGlyAlaAlaProGlyTyrThrThrIleGlyLeuTr 109  
 |||||  
 QY 896 GGAATGGATTTTA----- 908  
 |||||  
 Db 109 pAlaPheTyrLeuLeuIleValGlyArgValLeuValGlyLeuGlyValGlyAlaSe 129  
 |||||  
 QY 909 -----AAGGAGGTATGCCTTT 925

Db 129 rValLeuValProMetTyrIleSerGluIleAlaProIlyrAlaLeuArgGlyAlaLeuGl 149  
 |||||  
 QY 926 ACCTTTAATCCAAAGACGATCGCATCAGGATGTTGCTCCTCAATGCGTGGCAAGTTGTG 985  
 |||||  
 Db 149 ySerLeuTyrGlnLeuAlaIleThrIleGlyIleLeuValAlaAlaIleIleGlyLeuGl 169  
 |||||  
 QY 986 CATAACTACAGGAAG----- 1001  
 |||||  
 Db 169 yLeuAsnLysThrAsnAsnAspSerAlaLeuAsnSerTrpGlyTrpArgIleProLeuGl 189  
 |||||  
 QY 1002 -----CTGGTCTTTCTCTAGAGACTTCCTGATTCAAA 1033  
 |||||  
 Db 189 yLeuGlnLeuValProAlaLeuLeuLeuIleGlyLeuLeuPheLeuProGluSerP 209  
 |||||  
 QY 1034 ACATTTGGATCTTCAGAAAAGATCAAAATCTATTATTAGTGATAAAGCTTTAAGCTATACT 1093  
 |||||  
 Db 209 roArgTrpLeuVal-GluLysGlyLysLeuGluAlaArgGluValLeuAlaLysLeu 228  
 |||||  
 QY 1094 AAG-----CAGATACGAGCTCTCTC----- 1114  
 |||||  
 Db 229 ArgGlyValGluAspValAspGlnGlnIleGlnIleLysAlaGluLeuGluAlaGly 248  
 |||||  
 QY 1115 -----CTAAATTAACGATCATGCTGCGGTTATTGTTGGGATGTTGTCAGAAAC 1165  
 |||||  
 Db 249 ValGluGluLysAlaGlyLysAlaSerTrp-----GlyGluLeuPheArgGly 265  
 |||||  
 QY 1166 AGGCTAGATTCTCTACGGAAACTCTAAACGAAACGAAATTTTGTCAAGGTAAAGC 1225  
 |||||  
 Db 266 ArgThrArg-----ProIlyValArg 272  
 |||||  
 QY 1226 AGCAACTATCGAAGAAATGCGTTATCAGCGACCATCGGGCTTTTAGAGCTTTGGCA 1285  
 |||||  
 Db 273 GlnArgLeuLeuMetGlyValMetLeuGlnAlaPheGlnGlnLeuThrGlyIleAsnAla 292  
 |||||  
 QY 1286 ATCATC-----TTGCTCTATGTCAGTTTGGCTTTGATGCGATATGCT--- 1330  
 |||||  
 Db 293 IlePheTyrTyrSerProThrIlePheLysSerValGlyValSerAspSerArgAlaSer 312  
 |||||  
 QY 1331 -----TTCAGTCGCTATGCGCTTTAATTCATGACCTTTTGGCTACCTGCTGAGTCTTG 1384  
 |||||  
 Db 313 LeuLeuValThrIleIleValGlyValValAsnPheValPheThrLeuValAlaLeuIle 332  
 |||||  
 QY 1385 TTTATACACATTTCTTTTGAAGAAATTCAAATAGATTTCAGGCCATTTGGTCTTTA 1444  
 |||||  
 Db 333 PheLeuValAspArgPheGlyArgProLeuLeuLeuGlyAlaAlaGlyMetAla 352  
 |||||  
 QY 1445 ATG-----ACTGTATTGGGTATTTCATTAACAATCTTTGATCATTTTTCATCGTATT 1498  
 |||||  
 Db 353 IleCysPheLeuIleLeuGlyAlaSerIleGlyValAlaLeuLeuLeuAsnLysPro 372  
 |||||  
 QY 1499 CGTGAA-----GATCGCAAGCGAACTGTTTACCCTCATGCTTTTGTAGTTAATCAT 1552  
 |||||  
 Db 373 LysAspProLeuSerLysAlaAlaGlyIleValAlaIleValPheIleLeuLeuPheIle 392  
 |||||  
 QY 1553 GCC-----CTTCAAAAGACGTTTCAGC 1573  
 |||||  
 Db 393 AlaPhePheAlaLeuGlyTrpGlyProIleProTrpValIleLeuSerGluLeuPhePro 412  
 |||||  
 QY 1574 CGCAGGTAAATGCAACAGCTACCACTCATCATGTTTGTGTTTGTGTTTGTGTTTGTG 1627  
 |||||  
 Db 413 ThrIysValArgSerLysAlaLeuAlaLeuAlaThrAlaAlaAsnTrpLeuAlaAsn 432  
 |||||  
 QY 1628 ATAGCGGCTCTCTGCTCTTAAATTTTGCATTTTATTATG---ACCATAGGATCTTCTTA 1684  
 |||||  
 Db 433 IleIleGly-----PheLeuPheProTyrIleThrGlyAlaIleGlyLeuAlaLeu 449  
 |||||  
 QY 1685 GGAATCTTATGCTCTCTTATATTGACCACTCTCTGTTGTTTGTGTTTGTGTTTGTG 1735  
 |||||  
 Db 450 GlyGlyTyrValPheLeuValPheAlaGlyLeuLeuValLeuPheIleLeu 466  
 |||||

RESULT 30

US-09-795-693-27





```

Db 67 -----LeuLeuThrIleTyrPheLeu 74
QY 407 -----CTACAGTATTGGATGGCCACTCACCCTGTCAGACTGCTGG 451
Db 75 LysPheHisArgProMetTyrTyrPheIleGlyAsnLeuAlaLeuSerAspLeuAla 94
QY 452 ATTGTTCTTGCTATGGGATGGCCCTAGATGCAAAATGTTCTTGATTCGAAAGAATCCGA 511
Db 95 -----GlyValAlaIleTyrThrAlaAsnLeuLeu----- 104
QY 512 GAGGAATTTTATTGTCTCAAAAGTCTTAAAAATCTGAGAAAAAGATATACCAAGGCT 571
Db 104 ----- 104
QY 572 TTGGAGCCATTTTGGATCTAATCTAAGTACAGATATGGCCCTGACGCTTTCTTC 631
Db 105 -----SerGlyAlaThrTyrTyrLeuThrProAlaGlnTyrPhe 118
QY 632 CTAGATACAGGCGCTATTAAAGGTTTGGCTTGCATGATTGTAAGAAATTTCTCTTCA 691
Db 119 LeuArgGly-----SerMetPheValAlaLeuSerAlaSer 131
QY 692 ATGTTTACG-----GCTCTTTTCATGACTAATTTTCTTCATGCTGTGATG----- 739
Db 132 ValPheSerLeuLeuAlaIleAlaIleGlnArgTyrIleThrMetLeuLysMetLysLeu 151
QY 740 ---AATAGACCAACATACACAGTTGATGATGAAATTAAGTTGCTGGGATAAAGCAT 796
Db 152 HisAsnGlySerAsnAsnPheArgLeuLeuLeu----- 163
QY 797 GATTCTTGAAGAGATGCAAAAAAATTGGGCTGTTCTGCAAGCTTTTCTTTAGT 856
Db 164 -----IleSerAlaCys-----TyrValIleSer-----LeuIleLeuGly 175
QY 857 TGGCTGTGCTCGGGTTGGAGCCTGGAATCCGTTTGGGATGATGATTTAAAGAGAGG 916
Db 176 -----GlyLeuProIleMetGlyTyrAsnCysIle----- 185
QY 917 TATGCTTTTACCTTTAATCAAAAGAGATGGCATGCGATGTTCTCAAAATGCGTGGC 976
Db 185 ----- 185
QY 977 AAAGTTGCACTAAACTACAGAACTGCTTTTCTTGAAGACTTCGCTATTCAACA 1036
Db 186 -----SerAlaLeuSerSer----- 190
QY 1037 TTGGATCTTCAGAAAAAGATCAAAATCTAATTTAGTAAAGCTTTAAGCTATCTAAG 1096
Db 191 -----CysSerThrValLeuProLeuTyrHisLysHisTyrIleLeuPheCysThrThr 208
QY 1097 CAGATACAGAGCCTCTCTCTAATAATTAAAGATCATGAGCTGGCGTTATTGTGGGATT-- 1153
Db 209 ValPheThrLeuLeuLeuSerIleValIleLeu-----TyrCysArgLysIleTyr 225
QY 1154 ---GTTGTCAGAAACAGCCTAGATTTTCTTACGSAATCTTAAACGAACGAATAATTT 1210
Db 226 SerLeuValArgThrArgSerArgArgLeu-----ThrPheArgLysAsn----- 240
QY 1211 TGGTCAAAAGTAAAGAGCAAACTATCGAAGAAATGCTTATCAAGGACCATCGGCGCTT 1270
Db 241 -----IleSerLysAlaSerArgSerSerGlnAsnValAlaLeu 253
QY 1271 TTAGAGCTTTGGCAATCATCTTG---CTCATGTGAGTTTGGCGTTGAATGGCAATAT 1327
Db 254 LeuLysThrValIleIleValLeuSerValPheIleAlaCys-----TyrAlaPro 270
QY 1328 GCTTCACTGCCGTAATGCGCTTAAATCATGACCTTTGGCTACTGCTGCACTTCTTTT 1387
Db 271 LeuPheIleLeuLeuLeuAspValGlyCysLysValLysThrCysAspIleLeuPhe 290
QY 1388 ATAGCACTTTCTTTTGAAGAAATTCAAATAGATTTGCAAGCATTTGGGCTTATATG 1447

```

```

Db 291 ArgAlaGlyTyrPheLeu-----ValLeu 298
QY 1448 ACTGTATTGGGGATTTATTATTAACAATACTTTGATCATTTTGTATGCTATTCGTAAGAT 1507
Db 299 AlaValLeuAsnSerGlyThrAsnProIleIle----- 309
QY 1508 CGCCAAAGCAACCGTTTACCCCTATGATGCTTTAGTATGATGATGCTTCAAAAGACG 1567
Db 310 -----TyrThrLeuThrAsnLysGlnMetArgAla 320
QY 1568 TTCAGCCGC 1576
Db 321 PheIleArg 323

RESULT 33
US-09-731-030A-17
; Sequence 17, Application US/09731030A
; Patent No. US20020142375A1
; GENERAL INFORMATION:
; APPLICANT: MUNROE, Donald G
; APPLICANT: GUPTA, Ashwani K.
; APPLICANT: ZASTAMNY, Roman L.
; TITLE OF INVENTION: MAMMALIAN EDG-7 RECEPTOR HOMOLOGS
; FILE REFERENCE: 8074-0015
; CURRENT APPLICATION NUMBER: US/09/731, 030A
; PRIOR FILING DATE: 1998-12-29
; PRIOR APPLICATION NUMBER: 60/070,184
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Mammalian EDG
US-09-731-030A-17

Alignment Scores:
Pred. No.: 0.0409 Length: 381
Score: 101.00 Matches: 83
Percent Similarity: 31.83% Conservative: 58
Best Local Similarity: 18.74% Mismatches: 116
Query Match: 3.05% Indels: 186
DB: Gaps: 22

US-09-868-987-1 (1-1864) x US-09-731-030A-17 (1-381)
QY 290 ATCTCACAGATCGTGTGGCAATGCTTATTGTTTGAAGCGTATATTATGATTT 349
Db 53 IleLeuIleCysCys-----PheIleIleLeuGlnAsnIlePheVal----- 66
QY 350 GAGGCGCTCATCGCTTGGAGCGTGTCTTCTGATCTTTGCTTATCTGGGAGCT-- 406
Db 67 -----LeuLeuThrIleTyrPheLeu 74
QY 407 -----CTACAGTATTGGATGGCCACTCACCCTGTCAGACTGCTGG 451
Db 75 LysPheHisArgProMetTyrTyrPheIleGlyAsnLeuAlaLeuSerAspLeuAla 94
QY 452 ATTGTTCTTGCTATGGGATGGCCCTAGATGCAAAATGTTCTTGATTCGAAAGAATCCGA 511
Db 95 -----GlyValAlaIleTyrThrAlaAsnLeuLeu----- 104
QY 512 GAGGAATTTTATTGTCTCAAAAGTCTTAAAAATCTGAGAAAAAGATATACCAAGGCT 571
Db 104 ----- 104
QY 572 TTGGAGCCATTTTGGATCTAATCTAAGTACAGATATGGCCCTGACGCTTTCTTC 631
Db 105 -----SerGlyAlaThrTyrTyrLeuThrProAlaGlnTyrPhe 118
QY 632 CTAGATACAGGCGCTATTAAAGGTTTGGCTTGCATGATTGTAAGAAATTTCTCTTCA 691

```











```

; APPLICANT: Chu, Peter
; APPLICANT: Pardo, Jorge
; APPLICANT: Li, Congfen
; APPLICANT: Zhao, Haoran
; APPLICANT: Jiang, Yingping
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: EDG: Modulators of Lymphocyte Activation and Migration
; FILE REFERENCE: 021044-0003100S
; CURRENT APPLICATION NUMBER: US/09/971,228
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/284,763
; PRIOR FILING DATE: 2001-04-18
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 13
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: mutant #1, C-terminally truncated variant of
; OTHER INFORMATION: human endothelial differentiation G-protein
; OTHER INFORMATION: coupled receptor (GPCR) 1 (BDGI)
US-09-971-228-13

Alignment Scores:
Pred. No.: 0.112 Length: 341
Score: 96.50 Matches: 83
Percent Similarity: 32.05% Conservative: 59
Best Local Similarity: 18.74% Mismatches: 116
Query Match: 2.91% Indels: 185
DB: Gaps: 22

US-09-868-987-1 (1-1864) x US-09-971-228-13 (1-341)
QY 290 ATCTCAGCATGCTGGCTGGCAATGCTTATTTGTTGATGCGGTATATATAGATTT 349
||| |||||
Db 53 lLeuileucys-----PheileileuGluAsnlePheVal----- 66
QY 350 GGAGGGCTCATCGTCTTCGGGAGCTGTTCTTGAATCTTTTGGCTTATCTGGCGACGT--- 406
||| |||||
Db 67 -----LeuLeuThrIleTrpLysThrLys 74
QY 407 -----CTACAGTATTTGGATGGCGCCACTCACCCTGTGCAGGACTCGCTGGG 451
: : : |||
Db 75 LysPheHieArgProMetTyrTyrPheileGlyAsnLeuAlaLeuSerAspLeuLeuAla 94
QY 452 ATTGTTCTTCCTATGGGATGCCGTAGATGCAATGTTCTGTGATTCGAAAGATCCGA 511
||| |||||
Db 95 -----GlyValAlaTyrThrAlaAsnLeuLeu----- 104
QY 512 GAGGAATTTTATTGTCTCAAGTCTTAAAAATCTGTAGAAAAAGGATATACCAAGGCT 571
||| |||
Db 104 ----- 104
QY 572 TTGGAGCCATTTTGATTCTTAACCTGACTACAGTATTGGCTCAGCACTCTTTTCTTC 631
||| |||||
Db 105 -----SerGlyAlaThrTyrLysLeuThrProAlaGlnTrpPhe 118
QY 632 CTAGATACAGGGCTATTAAAGGTTTGCTTGACATGATTAGGAAATTTCTCTTCA 691
||| |||
Db 119 LeuArgLugly-----SerMetPheValAlaLeuSerAlaSer 131
QY 692 ATGTTTACG-----GCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGATC----- 739
: : : |||
Db 132 ValPheSerLeuLeuAlaIleAlaIleGluArgTyrIleThrMetLeuLysMetLysLeu 151
QY 740 ---AATAAGACCCAACATACACAGTTTGCATATATGTAAGTAAGTTCTGGGGGATAAGCAT 796
||| : : : |||
Db 152 HieAsnGlySerAsnAsnPheArgLeuPheLeu----- 163
QY 797 GATTTCTTGAGGAGATGCAAAAAAATTTCGGCTGTTTCTGGAAAGTGTTTTCTTTT 856
: : : |||
Db 164 -----IleSerAlaCys-----TrpValIleSer-----LeuileuGly 175

```



Query Match: 2.91% Indels: 185  
DB: 10 Gaps: 22

US-09-868-987-1 (1-1864) x US-09-993-844-8 (1-362)

QY 290 ATCTCAGCATCTGGCTTGGCAATGCTTATTTGTTGATGAGCGTATATATAGATTT 349  
DB 53 lLeuileCysCys-----PheileleLeuGluAsnilePheVal----- 66

QY 350 GGAGCGTCATCGCTTCGGGAGCTGTTCTTCTGAATCTTTTCTGCTATCTGGCGAGCT--- 406  
DB 67 -----LeuLeuThrlleTrpLysThrls 74

QY 407 -----CTACAGTATTGGATGCGCCACTCTGTCTAGGACTCGCTGGG 451  
DB 75 LysPheHisArgPrometTyTrpPheileGlyAsnLeuAlaLeuSerAspLeuLeuAla 94

QY 452 ATTGTTCTTGCTATGCGGATGCGCTAGATGCAATGTTCTTGTATTGCAAGAAATCCGA 511  
DB 95 -----GlyValAlaTyTrpAlaAsnLeuLeu----- 104

QY 512 GAGGAATTTTATTGCTCAAAAGTCTTAAAAAATCTGAGAAAAAGGATATACCAAGGCT 571  
DB 104 ----- 104

QY 572 TTGGAGCAATTTTGATCTTAACATTTGACTACAGTATTGGCTCAGCACTTCTTTTCTTC 631  
DB 105 -----SerGlyAlaThrTrpLysLeuThrProAlaGlnTrpPhe 118

QY 632 CTAGATACAGGCGCTATTAAAGGTTTCTTGTGACATTTAGTATTAGGAATTTCTCTTCA 691  
DB 119 LeuArgGluGly-----SerMetPheValAlaLeuSerAlaSer 131

QY 692 ATGTTTACG-----GCTCTTTTCATGACTAAATTTTCTCATCTCTGATG----- 739  
DB 132 ValPheSerLeuLeuAlaileGluArgTrpLleThrMetLeuLysMetLysLeu 151

QY 740 ---AATAAGACCAATACACAGTTGATATGATGAATAAGTTCTGCTGGGATAAAGCAT 796  
DB 152 HisAsnGlySerAsnAsnPheArgLeuPheLeuLeu----- 163

QY 797 GATTTCTTGAGAGGATGCAAAAACCTTTGGGCTGTTTCTGGNAGTGTCTTTCTTTAGGT 856  
DB 164 -----IleSerAlaCys-----TrpValIleSer-----LeuileLeuGly 175

QY 857 TCGGTTGCTCTCGGTTGGAGCTGGAATCCGTTTGGGAATGCAATTTAAAGGAGGG 916  
DB 176 -----GlyLeuProIleMetGlyTrpAsnCysIle----- 185

QY 917 TATGCTTTTACCTTTAATCCAAAAGCATGGCATCAGCGATGTTTGCTCAAAATGCTGGC 976  
DB 185 ----- 185

QY 977 AAAGTTGTCATAAACTACAGAACTGGTCTTTCTTCTAGAGACTTCGGTATTCAAACA 1036  
DB 186 -----SerAlaLeuSerSer----- 190

QY 1037 TTGGATCTTCAGAAAAGATCAAAATCTATTATTAGTATAAAGCTTTAAAGCTATATAAG 1096  
DB 191 -----CysSerThrValLeuProLeuTyHisLysHisTyrlleLeuPheCysThrThr 208

QY 1097 CAGATACGAGCTCTCTCTAAATTAACGATCATGAGTGGCGTATTGTGGGATT--- 1153  
DB 209 ValPheThrLeuLeuLeuSerileValIleLeu-----TyCysArgileTyTr 225

QY 1154 ---GTTGTCAGAAACAGGCTAGATTTCTTACGGAATACTTAAACGAAAGCAAAATTT 1210  
DB 226 SerLeuValArgThrArgSerArgLeu-----ThrPheArgLysAsnileSerLys 243

QY 1211 TGGTCAAAAGGTAAGCAGCAAACTATCGAAGAAATCGGTTATCAGCGCACCATCGGCTT 1270  
DB 244 AlaSerArgSerSerGluLys-----SerLeuAlaLeu 254

QY 1271 TTAGGAGCTTTGGCAATCATCTTG---CTCTATGTAGTTTGGCTTTGAATGGCAATAT 1327  
DB 255 LeuLysThrValIleleValLeuSerValPheileAlaCys-----TrpAlaPro 271

QY 1328 GCTTTAGTCCGCTATCGCTTTAAATTCATGACCTTTTGGCTACCTGTGAGTCTTCTTT 1387  
DB 272 LeuPheileleLeuLeuLeuAspValGlyCysLysValLysThrCysAspIleLeuPhe 291

QY 1388 ATAGCACATTTCTTTTGAAGAAAATTCAAATAGATTTGCAAGCCATTGTGCTTTAATG 1447  
DB 292 ArgAlaGluTyTrpPheLeu-----Valleu 299

QY 1448 ACTGATTGGGTATTTCATTAACAATACCTTTTGATCATTTTGTGATCGTATTCTGGAAGAT 1507  
DB 300 AlaValLeuAsnSerGlyThrAsnProIleile----- 310

QY 1508 CGCCAAGCGAACCTGTTTACCCTCATGCTTTTACTGATTAATGATGCCCTTCAAAGACG 1567  
DB 311 -----TyThrLeuThrAsnLysGluMetArgArgAla 321

QY 1568 TTCAGCCGC 1576  
DB 322 PheileArg 324

RESULT 41  
US-10-037-616-19  
; Sequence 19, Application US/10037616  
; Patent No. US20020123148A1  
; GENERAL INFORMATION:  
; APPLICANT: English, Denis  
; APPLICANT: Kovacs, Richard J.  
; APPLICANT: Rizzo, Maria T.  
; APPLICANT: Sliva, Daniel T.  
; TITLE OF INVENTION: Sphingolipid Compositions and Methods for Their Therapeutic Use  
; FILE REFERENCE: 7042-119  
; CURRENT APPLICATION NUMBER: US/10/037,616  
; CURRENT FILING DATE: 2002-04-19  
; PRIOR APPLICATION NUMBER: US 60/243,887  
; PRIOR FILING DATE: 2000-10-27  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 19  
; LENGTH: 381  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-037-616-19

Alignment Scores:  
Pred. No.: 0.117 Length: 381  
Score: 96.50 Matches: 83  
Percent Similarity: 32.05% Conservative: 59  
Best Local Similarity: 18.74% Mismatches: 116  
Query Match: 2.91% Indels: 185  
DB: 12 Gaps: 22

US-09-868-987-1 (1-1864) x US-10-037-616-19 (1-381)

QY 290 ATCTCAGCATCTGGCTTGGCAATGCTTATTTGTTGATGAGCGTATATATAGATTT 349  
DB 53 lLeuileCysCys-----PheileleLeuGluAsnilePheVal----- 66

QY 350 GGAGCGTCATCGCTTCGGGAGCTGTTCTTCTGAATCTTTTGTCTTATCTGGCGAGCT--- 406  
DB 67 -----LeuLeuThrlleTrpLysThrls 74

QY 407 -----CTACAGTATTGGATGCGCCACTCTGTCTAGGACTCGCTGGG 451  
DB 75 LysPheHisArgPrometTyTrpPheileGlyAsnLeuAlaLeuSerAspLeuLeuAla 94

QY 452 ATTGTTCTTGCTATGCGGATGCGCTAGATGCAATGTTCTTGTATTGCAAGAAATCCGA 511  
DB 95 -----GlyValAlaTyTrpAlaAsnLeuLeu----- 104

QY 512 GAGGAATTTTATGTCTCAAGTCTTAATAAATCTGTAGAAAAGATATACCAAGCT 571  
 Db 104 ----- 104  
 QY 572 TTTGAGCCATTTTGAATCTAAGTCTGACATATGGCCCTGAGCATTTCTTTCTTC 631  
 Db 105 -----SerGlyAlaThrThrIleuThrIleuThrProIleuThrPhe 118  
 QY 632 CTAGATACAGGCGCTTAAGAGGTTGGCTTGAACATGATTTTGAAGATTTCTCTTCA 691  
 Db 119 LeuArgIleuGly-----SerMetPheValAlaIleuSerAlaSer 131  
 QY 692 ATGTTTACG-----GCTCTTTTCATGACTAATTTTCTTCATGCTGTGAGT----- 739  
 Db 132 ValPheSerIleuLeuAlaIleuAlaIleuIleuArgIleuThrMetLeuLysMetLysLeu 151  
 QY 740 ----AATAAGCCCAACATACAGTGTGATGATGATAGTGTGGGATTAACAT 796  
 Db 152 HisasnGlySerAsnAsnPheArgLeuPheLeu----- 163  
 QY 797 GATTTCCTGAGAGATGCAAAAACTTGGCGTGTCTGAGAGTGTCTTTCTTTAGT 856  
 Db 164 ----IleSerAlaCys-----TryValIleSer-----LeuIleLeuGly 175  
 QY 857 TGCCTTGTCTCGGCTTGGAGCTGGAATCCGTTTGGGAATGATTTTAAAGAGG 916  
 Db 176 ----GlyLeuProIleMetGlyTryPasnCysIle----- 185  
 QY 917 TATGCTTACCTTAAATCCAAAAGAGATGACATGAGATGTGCTCAATGCGTGGC 976  
 Db 185 ----- 185  
 QY 977 AAAGTTGTCAATACTACAGAAAGCTGTCTTCTTCTAGACCTCCGATTCAACA 1036  
 Db 186 -----SerAlaLeuSer----- 190  
 QY 1037 TTTGATCTTCAAGAAAGATCAAAATCTATTACTGATTAAGCTTAAAGCTATATAG 1096  
 Db 191 ----CysSerThrValLeuProLeuThrHisLysHisTyrIleLeuPheCysThrThr 208  
 QY 1097 CAGATACAGGCTCTCTCTAAATTAACATCAATGAGCTGGCGTTATTTGGAGT--- 1153  
 Db 209 ValPheThrIleuLeuLeuSerIleValIleLeu-----TyrCysArgIleTyr 225  
 QY 1154 ----GTTCAGAAAACAGCCCTAGATTCTTACAGGAACTCTAAACGAAACGAAATTT 1210  
 Db 226 SerIleuValArgThrArgSerArgArgLeu-----ThrPheArgLysAsnIleSerLys 243  
 QY 1211 TGGTCAAAAGGTACAGCAACTATCGAAGAAATGCTTACAGGCGACCATGGGCTT 1270  
 Db 244 AlaSerArgSerSerGlyLys-----SerLeuAlaLeu 254  
 QY 1271 TTAGAGCTTTGGCAATCATCTTG--CTATATGAGTTTGGCTTTGAATGCAATAT 1327  
 Db 255 LeuIleThrValIleIleValLeuSerValPheIleAlaCys-----TyrAlaPro 271  
 QY 1328 GCTTTCAGTCCGTATGCGCTTAAATTCATGACCTTTTGGTACCTGTGAGCTTGT 1387  
 Db 272 LeuIleThrIleuLeuLeuAspValGlyCysLysValLysThrCysAspIleLeuPhe 291  
 QY 1388 ATGACACTTTCTTTTGAAGAAAATTCAAATAGTTTGCAGAGCATGTGGCTTTATG 1447  
 Db 292 ArgIleGlyIleThrPheLeu-----ValLeu 299  
 QY 1448 ACTGATTTGGGGTATTCATTAACAATATCTTGATCATTTTGTATCGTATCGTAAGAT 1507  
 Db 300 AlaValLeuAsnSerGlyThrAsnProIleIle----- 310  
 QY 1508 CGCCAGAGCAACTGTATACCCCTATGATGATTTTGAATGATGATGATGATGATGATG 1567  
 Db 311 ----TyrThrIleuThrIleuLysGluMetArgArgAla 321  
 QY 1568 TTCAGCCGC 1576

Db 322 PheIleArg 324  
 RESULT 42  
 US-09-971-228-5  
 / Sequence 5, Application US/09971228  
 / Patent No. US2002015512A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Liao, X. Charlene  
 / APPLICANT: Masuda, Esteban  
 / APPLICANT: Chu, Peter  
 / APPLICANT: Pardo, Jorge  
 / APPLICANT: Li, Congfen  
 / APPLICANT: Zhao, Haoran  
 / APPLICANT: Jiang, Yingping  
 / APPLICANT: Rigel Pharmaceuticals, Incorporated  
 / TITLE OF INVENTION: EDG: Modulators of Lymphocyte Activation and Migration  
 / FILE REFERENCE: 021044-000310US  
 / CURRENT APPLICATION NUMBER: US/09/971,228  
 / CURRENT FILING DATE: 2002-04-19  
 / PRIOR APPLICATION NUMBER: US 60/284,763  
 / PRIOR FILING DATE: 2001-04-18  
 / NUMBER OF SEQ ID NOS: 15  
 / SOFTWARE: PatentIn Ver. 2.1  
 / SEQ ID NO 5  
 / LENGTH: 382  
 / TYPE: PRT  
 / ORGANISM: Homo sapiens  
 / FEATURE:  
 / OTHER INFORMATION: wild-type human endothelial differentiation  
 / OTHER INFORMATION: G-protein coupled receptor (GPCR) 1 (EDG1)  
 US-09-971-228-5

Alignment Scores:  
 Pred. No.: 0.117 Length: 382  
 Score: 96.50 Matches: 83  
 Percent Similarity: 32.05% Conservative: 59  
 Best Local Similarity: 18.74% Mismatches: 116  
 Query Match: 2.91% Indels: 185  
 DB: Gaps: 22  
 US-09-868-987-1 (1-1864) x US-09-971-228-5 (1-382)  
 QY 290 ATCTCAGATGCTGTGGCTTGGCAATGCTTATTTGATGAGCTATATTAAGATT 349  
 Db 53 IleLeuIleCysCys-----PheIleIleLeuGluAsnIlePheVal----- 66  
 QY 350 GAGAGGCTATCGCTTCGGAGCGTTCCTGATCTTTTGTCTTACGCGAGCT--- 406  
 Db 67 -----LeuLeuThrIleTyrPheLys 74  
 QY 407 -----CTACATATTTGATGATGCGCACTCACTTGTGAGACTCGCTGG 451  
 Db 75 LysPheHisArgProMetTyrTyrPheIleGlyAsnLeuAlaLeuSerAspLeuAla 94  
 QY 452 ATTGTTCTTCTATGCGGATGCGCTAGATGCAAAATTTTGTATTCGAAAGATCCGA 511  
 Db 95 -----GlyValAlaTyrThrAlaAsnLeuLeu----- 104  
 QY 512 GAGGAATTTTATGTCTCAAGTCTTAATAAATCTGTAGAAAAGATATACCAAGCT 571  
 Db 104 ----- 104  
 QY 572 TTTGAGCCATTTTGAATCTAAGTCTGACATATGGCCCTGAGCATTTCTTTCTTC 631  
 Db 105 -----SerGlyAlaThrThrIleuThrIleuThrProIleuThrPhe 118  
 QY 632 CTAGATACAGGCGCTTAAGAGGTTGGCTTGAACATGATTTTGAAGATTTCTCTTCA 691  
 Db 119 LeuArgIleuGly-----SerMetPheValAlaIleuSerAlaSer 131  
 QY 692 ATGTTTACG-----GCTCTTTTCATGACTAATTTTCTTCATGCTGTGAGT----- 739



QY 1115 -----CTAAATTACGATCATGAGCTGGCCTTATTTGGGATTTGTTCAGAAAC 1165  
 Db 249 ValIugluGluLysAlaGlyLysAlaSerTyr-----GlyGluLeuPheArgGly 265  
 QY 1166 AGGCTAGATTCTCTACGAAACTCTAAACGAAACGAAATTTGGTCAAGGTAAAGC 1225  
 Db 266 ArgThrArg-----ProLysValArg 272  
 QY 1226 AGCAACTGATCGAAGAAATGCGTTATGAG----- 1255  
 Db 273 GlnArgLeuLeuMetGlyValMetLeuGlnPheGlnLeuThrGlyIleAsnAlaIle 292  
 QY 1256 -----GCGACATCGGGCTTTTGAAGAGCTTTGGCATATC 1291  
 Db 293 PheTyrTyrSerProThrIlePheLysSerValGlyValSerAspSerArgLaserIleu 312  
 QY 1292 TTGGCTATGATGATTTGGCTTGAATGCAATATGCTTTCAGTGGCGTATGCGCTTTA 1351  
 Db 313 LeuValThrIle-----IleValGlyVal 320  
 QY 1352 ATTGATGACCTTTGGCTACCTGCTGCTGCTTTTATAGACATTTCTTTTGAAGAA 1411  
 Db 321 ValAsnPheValPheThrLeuValAlaLeuIlePheLeuValAspArgPheGlyArgArg 340  
 QY 1412 ATTCAATAGATTGCAAGCCATTTGGCTTATG-----ACTGATTTGGGCTATTC 1465  
 Db 341 ProLeuLeuLeuGlyAlaIleGlyMetAlaIleCysPheLeuIleLeuGlyAlaSer 360  
 QY 1466 TTAACAATACCTTTGATTTTGTATGATTCGTATGCTGAA-----GATCGCAAGCGAAC 1519  
 Db 361 IleGlyValAlaLeuLeuLeuLeuAsnLysProLysAspProLeuSerLysAlaIleGly 380  
 QY 1520 CTGTTTACCCCTATGATGATTTTATGATGATGCT----- 1555  
 Db 381 IleValAlaIleValPheIleLeuPheIleAlaPhePheAlaLeuGlyTyrGlyPro 400  
 QY 1556 -----CTTCAAAAGACGTTTCAGCCGACGCTATGACACAGCTACCACT 1600  
 Db 401 IleProTyrValIleLeuSerGluLeuPheProThrLysValArgSerLysAlaLeuAla 420  
 QY 1601 CTATCAGTTTGTATATGCTTTTG-----TTTATAGCGGCTCCTGCTTTTATATTT 1654  
 Db 421 LeuAlaThrAlaAlaAsnTyrLeuAlaAsnPheIleIleGly-----PheLeuPhe 437  
 QY 1655 GCATTTATTATG-----ACCATAGGAGATTTCTTACGAACTTATGCTCTTTTATATGCA 1711  
 Db 438 ProTyrIleThrGlyAlaIleGlyLeuAlaLeuGlyGlyTyrValPheLeuValPheAla 457  
 QY 1712 CCACCTCTGTTGTTTATGATGCTC 1735  
 Db 458 GlyLeuLeuValLeuPheIleLeu 465  
 RESULT 44  
 ; Sequence 4, Application US/09741153  
 ; Patent No. US20020102637A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YAN, Chunhua et al  
 ; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
 ; TITLE OF INVENTION: AND USES THEREOF  
 ; FILE REFERENCE: C1001059  
 ; CURRENT APPLICATION NUMBER: US/09/741,153  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 619  
 ; TYPE: PRT  
 ; ORGANISM: Rat  
 ; US-09-741-153-4

## Alignment Scores:

Pred. No.:	0.162	Length:	619
Score:	96.00	Matches:	117
Percent Similarity:	33.33%	Conservative:	77
Best Local Similarity:	20.10%	Mismatches:	192
Query Match:	2.90%	Indels:	196
DB:	10	Gaps:	30

US-09-868-987-1 (1-1864) x US-09-741-153-4 (1-619)

QY 248 ATCTCTTCATGATCTTGGGAAAAAACAATGATACAGAGCATTAATCTCAGCATCTGGC 307  
 Db 53 LeuAlaGlyGluValAlaLysGluLysAlaGlyProSerIleValIleCysPheLeuVal 72  
 QY 308 TTGGCAATGCTTATTTGTTGATGACCGTA---TATTATGATTTGGAGGC---GTCATC 361  
 Db 73 AlaIleLeuSerSerValLeuAlaGlyLeuCysTyrAlaGlyIlePheGlyAlaArgValPro 92  
 QY 362 GCTTGGGAGCTGTTCTTGTGAATCTTTGCTT-----ATCTGGGCA----- 403  
 Db 93 GlySerLysSerAlaTyrLeuTyrSerTyrValThrValGlyGluLeuTyrPalaPheThr 112  
 QY 404 -----GCTCTACAGTATTTGATGACCGCACTGCTGCTGACAGACTCGCT 448  
 Db 113 ThrGlyTyrAsnLeuIleLeuSerTyr-----ValIleGlyThrAla 126  
 QY 449 GGGATTTGTTCTTCTATGCGGATGCGCTGATGCAATGTTCTGTATTCGAAAGATC 508  
 Db 127 SerValAlaArgAlaTyrSerSerAlaPheAspAsn-----LeuIleGlyAsnHisIle 144  
 QY 509 CGAGAGGAATTTTATTTGCTCAAGCTTTAAATAATCTGTAGAAAGAGATATACCAAG 568  
 Db 145 -----SerGlnIleLeuLys----- 149  
 QY 569 GCTTTTGAAGCCATTTTGTATTTACTTACTGATGATGATGAGCTTTCAGACTTTTTC 628  
 Db 150 -----GlyThrIleLeu-----LeuAsnMetProHisValLeuAlaGlyLysTyrProAspPhe 166  
 QY 629 TTCTTATGATACAGGCGCTTATTAAGGCTTTCCTTACATGATTTTA----- 676  
 Db 167 Phe-----AlaLeuAlaLeuValLeuLeuLeuThrGlyLeuLeu 179  
 QY 677 -----GGAATTTCTCTGCAATGTTTACAGGCTCTTTTACGACT 715  
 Db 180 ValLeuGlyAlaAsnGluSerGlyLeuValThrLysValPheThrGlyMetAsnLeuLeu 199  
 QY 716 AATTTTCTTCAATGCTGTGATGATGATGATGACCAACATACACAGTGCATATGATGAAT 775  
 Db 200 ValLeuGlyPheVal----- 204  
 QY 776 AAGTTGTTGGGATTAAGCATGATTTCTTGAGAGATGCAAAAACTTTGGGCTGTTCT 835  
 Db 205 -----IleIleSerGlyPheIleLysGlyGluLeuArgAsnTyrLysLeuThr 220  
 QY 836 GGAAGTGTCTTTCTTTAGTTGCGTTCGCTTCGGGTTTGGAGCCGTGAATCCGTTTG 895  
 Db 221 LysGluAspTyr-----CysLeuThrMetSerGluSer-----AsnGlyThrCys 235  
 QY 896 GGAATGATTTTAAAGAGAGGATATGCTTTTACCTTTTAAATCCAAAAGAGATGATCAGC 955  
 Db 236 SerLeuAspSerMetCysGlySerGlyPhe-----MetProPheGlyLeuGly 251  
 QY 956 GATGTGCTCAAAATGCGTGGCAAGTTGTCATAACTACAGAGAGCTGCTTTCTTCT 1015  
 Db 252 GlyIle-----LeuArgGlyAlaAlaThrCysPheTyrAlaPheValGlyPheAspCys 269  
 QY 1016 AGAGACTTCGATTTCAACATTTTGATCTTCGAGAAAGATCAAAATCTATTTTATGAT 1075  
 Db 270 -----IleAlaThrThrGly-----Glu 275  
 QY 1076 AAGCTTTAAGCTATCTACAGAGATACAGAGCTCTCTC---CTAAATTTAAGATCAG 1132  
 Db 276 GluAlaGlnAsnProGlnIleArgSerIleProMetGlyIleValIleSerLeuSerIleCys 295



```
QY 1133 ACCTGGCGTTATTGTGGGATTTGTTGTGAGAAACAGCGCTAGATTCTCTACGGAAACTCT 1192
Db 296 PheLeuAlaTyrPheGly-----
QY 1193 AAACGAAACGCAAAATTTTGGTCAAAGGTAAGCAGCAAACTATCGAAGAAATG----- 1246
Db 302 -----ValSerSerAlaLeuThrLeuMetMetProTyr 312
QY 1246 ----- 1246
Db 313 TyrLysLeuGlnProGluSerProLeuProGluAlaPheThrTyrValGlyTrpGluPro 332
QY 1247 ----CGTTATCAGCGACACCTCGGCTTTTGGAGCTTTGGCAATCATCTTGCTC----- 1297
Db 333 AlaArgTyrLeuValAlaIleGlySerLeuCysAlaLeuSerThrSerLeuLeuGlySer 352
QY 1298 ---TATGTGAGTTTGGCTTTGAATGGCAATATGCT-----TTCAAGTCCGCTA 1342
Db 353 MetPheProMetProArgValIleTyrAlaMetAlaGluAspGlyLeuLeuPheArgVal 372
QY 1343 TCGCGCTTAATTATCATGAC-----CTTTTGGCTACCTGTGCA----- 1378
Db 373 LeuAlaArgValHisAsnGlyThrHisThrProIleValAlaThrValValSerGlyVal 392
QY 1379 ---GTCTTGTATTATAGCAATTTCTTTTGAAGAAATTCAAATAGATTTGCAAGCAATT 1435
Db 393 IleAlaAlaPheMetAlaPheLeuPheGluLeuThrAspLeuValAspLeuMetSerIle 412
QY 1436 GTGCTCTTAATGACTGTATTGGGTATTATCATTAACAATCTTTGATCATTTTGTATCGT 1495
Db 413 Gly-----ThrLeuAlaTyrSerLeuValSerIleCysValleuIleLeuArg 429
QY 1496 ATTCGTGAAGATCGC----- 1510
Db 430 TyrGlnProAspGlnGluMetLysAsnGlyGluGluValGluLeuGlnGluArg 449
QY 1511 -----CAAGCGAAGCTGTTTACCCTATGCAATGTTTATAGTTAAT---GATCCCTTCAA 1561
Db 450 ThrLeuGluAlaGluLysLeuThrValGlnAlaLeuPheCysGlnValAspSerIlePro 469
QY 1562 AAGAGCTTCAGC---CGCAGCGTAATGACACAGCTACAACTCTATCAGTTTGTGTAATG 1618
Db 470 ThrLeuLeuSerGlyArgIleValTyrValCysSerSerLeuAlaValleuLeuThr 489
QY 1619 CTTTGTGTTTATA-----GGCGGCTCCTCTGCTTT 1648
Db 490 ValLeuCysLeuValLeuThrTrpTrpThrProLeuHisSerGlyAspProValTrp 509
QY 1649 AATTTGCAATTTATTATGACCATAGGATTTCTTCTAGGAATTTATCGTCTCTTTATATT 1708
Db 510 ValThrValValleuLeuLeuGlyLeuIleLeuGlyIleSerGlyValIleTrpArg 529
QY 1709 GCACCA 1714
Db 530 GlnPro 531

RESULT 45
US-09-971-228-14
; Sequence 14, Application US/09971228
; Patent No. US2002015512A1
; GENERAL INFORMATION:
; APPLICANT: Liao, X. Charlene
; APPLICANT: Masuda, Eateban
; APPLICANT: Chu, Peter
; APPLICANT: Pardo, Jorge
; APPLICANT: Li, Congfen
; APPLICANT: Zhao, Haoan
; APPLICANT: Jiang, Yingping
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: EDG: Modulators of Lymphocyte Activation and Migration
; FILE REFERENCE: 021044-000310US
; CURRENT APPLICATION NUMBER: US/09/971,228
```

```
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/284,763
; PRIOR FILING DATE: 2001-04-18
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: mutant #2, C-terminally truncated variant of
; OTHER INFORMATION: human endothelial differentiation G-protein
; OTHER INFORMATION: coupled receptor (GPCR) 1 (EDG1)
US-09-971-228-14
Alignment Scores:
Pred. No.: 0.173 Length: 314
Score: 94.50 Matches: 76
Percent Similarity: 33.16% Conservative: 52
Best Local Similarity: 19.69% Mismatches: 103
Query Match: 2.85% Indels: 155
DB: 9 Gaps: 20
```

```
US-09-868-987-1 (1-1864) x US-09-971-228-14 (1-314)
QY 290 ATCTCAGCATCTGTGGCTTGGCAATGCTTATGTTTGTATGATGAGCGTATATTATAGATT 349
Db 53 IleLeuIleCysCys-----PheIleIleLeuGluAsnIlePheVal----- 66
QY 350 GGAGCGTCATCGCTCGGGAGCTGTTCTTCTGAATCTTTTGTCTATCTGGCAGCT--- 406
Db 67 -----LeuLeuThrIleTrpLysThrLys 74
QY 407 -----CTACAGTATTTGGATGGCCACCTGCTGACAGGACTCGCTGG 451
Db 75 LysPheHisArgProMetTyrTyrPheIleGlyAsnLeuAlaLeuSerAspLeuLeuAla 94
QY 452 ATGTGTTCTGTCTGGGATGCGCCGTAGATGCAAAATGTTCTTGTATTTCGAAAGAAATCCGA 511
Db 95 -----GlyValAlaTyrThrAlaAsnLeuLeuLeu----- 104
QY 512 GAGGAATTTTATTGTCTCAAAGTCTTAAANAATCTGTAGAAAAGGATATACCAGGCT 571
Db 104 ----- 104
QY 572 TTTGAGGCCATTTTGTATTCTAATCTGACTACAGTATTGGCTCAGCACTTCTTTTCTTC 631
Db 105 -----SerGlyAlaThrThrTyrLysLeuThrProAlaGlnTrpPhe 118
QY 632 CTAGATACAGGCGCTATTAAAGGGTTTGTCTTGACATTGATTTTAGGAATTTTCTCTTCA 691
Db 119 LeuArgGluGly-----SerMetPheValAlaLeuSerAlaSer 131
QY 692 ATGTTTAGC-----GCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGATG----- 739
Db 132 ValPheSerLeuLeuAlaIleAlaIleGluArgTyrIleThrMetLeuLysMetLysLeu 151
QY 740 ---AATAAGACCCACATACACAGTTGCATATGATGAATAAGTTCTGCGGGATAAAGCAT 796
Db 152 HisAsnGlySerAsnAsnPheArgLeuPheLeuLeu----- 163
QY 797 GATTTCTTGAGAGGATGCAAAAAAATTTTGGGCTGTTTCTGGAAGTGTTTTCTTTTAGGT 856
Db 164 -----IleSerAlaCys-----TrpValIleSer-----LeuIleLeuGly 175
QY 857 TCGGTGCTCTCGGGTTTGGAGCCTGGAATTCGGTTTGGGAATGGATTTTAAAGGAGGG 916
Db 176 -----GlyLeuProIleMetGlyTrpAsnCysIle----- 185
QY 917 TATGCTTTACTTTAATCCAAAAGAGCATGGCATCAGCGATGTTGTCTCAATGCGTGGC 976
Db 185 ----- 185
```

```
QY 977 AAAGTTGATTAACCTACGAGAGCTGGCTTTCTTACAGACTCCGATTCAACA 1036
Db 186 -----SerAlaLeuSer----- 190
QY 1037 TTGGATCTTCAGAAAAGATCAAAATCTATTATTAGATAAAGCTTAAGCTAATAAG 1096
Db 191 -----CysSerThrValLeuProLeuThrHisIysHisTyrIleLeuPheCysThrThr 208
QY 1097 CAGATACGAGCTCTCTCTTAAATTAACGATCATGAGCTGGCGTTATGTGGATT--- 1153
Db 209 ValPheThrLeuLeuLeuSerIleValIleLeu-----TyrCysArgIleTyr 225
QY 1154 ---GTTGTCAGAAACAGGCCCTAGATTCTCTACGGAACCTAAACGAACGCAAAATTT 1210
Db 226 SerLeuValArgThrArgSerArgArgLeu-----ThrPheArgLysAsnIleSerLys 243
QY 1211 TGGTCAAGGTAAAGCAGCAAACTATCGAAGAAATGCGTTATCAGCGACCATCGGCTT 1270
Db 244 AlaSerArgSerSerGluLys-----SerLeuAlaLeu 254
QY 1271 TTAGAGCTTTGGCAATCATCTTG--CTCATGTGAGTTTGGCTTTGAATGGCAATAT 1327
Db 255 LeuLysThrValIleIleValLeuSerValPheIleAlaCys-----TrpAlaPro 271
QY 1328 GCTTTCAGTCCGATATGCGCTTTAATCATGACCTTTTGCTACCTGTGCAGTCTTGT 1387
Db 272 LeuPheIleLeuLeuLeuLeuAspValGlyCysLysValIlyThrCysAspIleLeuPhe 291
QY 1388 ATAGCACATTTCTTTTG 1405
Db 292 ArgAlaGluTyrPheLeu 297
```

Search completed: December 17, 2002, 12:50:03  
Job time : 89.5 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2002, 06:25:53 ; Search time 1976 Seconds  
(without alignments)  
15277.524 Million cell updates/sec

Title: US-09-868-987-1  
Perfect score: 1864  
Sequence: 1 atgacttcgcattgtc.....atgacaattcagataatgc 1864

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estinv.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_htc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_htc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: gb\_gss.\*

18: em\_gss\_hum.\*

19: em\_gss\_inv.\*

20: em\_gss\_pln.\*

21: em\_gss\_vrt.\*

22: em\_gss\_fun.\*

23: em\_gss\_mam.\*

24: em\_gss\_mus.\*

25: em\_gss\_other.\*

26: em\_gss\_pro.\*

27: em\_gss\_rod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55.8	3.0	1101	17	CNS00LT2
2	54.4	2.9	187	17	BH371348
3	51.8	2.8	602	17	BH182894
4	51.8	2.8	602	17	CNS0708R
5	50.8	2.7	731	17	BH397261
6	49.2	2.6	1101	17	CNS00FOO

7	48	2.6	1101	17	CNS0039G
8	47.2	2.5	1101	17	CNS0100X
9	46.8	2.5	1101	17	CNS000D1
C 10	46.6	2.5	685	13	BJ353037
C 11	45.4	2.4	1101	17	CNS00807
C 12	45.2	2.4	963	9	AL566565
C 13	45.2	2.4	988	17	CNS06PBX
14	44.4	2.4	516	10	BE118484
15	43.8	2.3	548	17	AZ303132
C 16	43.8	2.3	1125	9	AL547503
17	43.6	2.3	417	14	BQ399330
18	43.6	2.3	500	9	AU087215
C 19	43.6	2.3	953	17	CNS02701
20	43.4	2.3	1101	17	CNS00HFX
C 21	43.2	2.3	803	17	CNS00DM2
C 22	43.2	2.3	863	17	AZ550729
C 23	42.8	2.3	332	12	BG350812
C 24	42.6	2.3	1001	17	CNS0155H
C 25	42.4	2.3	374	14	T72643
C 26	42.4	2.3	713	17	B80295
C 27	42.4	2.3	878	17	CNS0187R
C 28	42.4	2.3	885	17	CNS0187R
C 29	42.2	2.3	1101	17	CNS0039G
C 30	41.8	2.2	946	17	CNS005V6
31	41.8	2.2	976	17	CNS04ESM
C 32	41.8	2.2	1096	17	CNS051P2
C 33	41.8	2.2	1101	17	CNS003B4
C 34	41.6	2.2	930	17	AZ680882
35	41.6	2.2	931	17	AZ541040
36	41.6	2.2	1080	17	CNS008PP
37	41.4	2.2	552	12	BF279297
C 38	41.4	2.2	676	13	BJ349244
C 39	41.4	2.2	924	12	BG576544
40	41.4	2.2	1101	17	CNS008N7
C 41	41.4	2.2	1101	17	CNS001Q0
C 42	41.2	2.2	546	17	AZ036289
C 43	41.2	2.2	629	12	BF295860
C 44	41.2	2.2	750	17	CNS0111D
C 45	41	2.2	436	14	C92475

## ALIGNMENTS

CNS00LT2 1101 bp DNA linear GSS 14-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC:  
BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL078714 GI:5102004

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Genoscope.  
Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Ooeogawa and Aaron Mammose in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial

ECORI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

## Source

1. 1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BACR48P19"  
/clone\_lib="RPCI-98"  
/note="end : TERT"

BASE COUNT 469 a 6 c 69 g 151 t 406 others

## ORIGIN

Query Match 3.0%; Score 55.8; DB 17; Length 1101;  
Best Local Similarity 21.9%; Pred. No. 0.002;  
Matches 117; Conservative 165; Mismatches 253; Indels 0; Gaps 0;

## FEATURES

## Source

1. 1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BACR48P19"  
/clone\_lib="RPCI-98"  
/note="end : TERT"

BASE COUNT 469 a 6 c 69 g 151 t 406 others

## ORIGIN

Query Match 3.0%; Score 55.8; DB 17; Length 1101;  
Best Local Similarity 21.9%; Pred. No. 0.002;  
Matches 117; Conservative 165; Mismatches 253; Indels 0; Gaps 0;

## FEATURES

## Source

1. 1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BACR48P19"  
/clone\_lib="RPCI-98"  
/note="end : TERT"

BASE COUNT 469 a 6 c 69 g 151 t 406 others

## ORIGIN

Query Match 3.0%; Score 55.8; DB 17; Length 1101;  
Best Local Similarity 21.9%; Pred. No. 0.002;  
Matches 117; Conservative 165; Mismatches 253; Indels 0; Gaps 0;

## FEATURES

## Source

1. 1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BACR48P19"  
/clone\_lib="RPCI-98"  
/note="end : TERT"

BASE COUNT 469 a 6 c 69 g 151 t 406 others

## ORIGIN

Query Match 3.0%; Score 55.8; DB 17; Length 1101;  
Best Local Similarity 21.9%; Pred. No. 0.002;  
Matches 117; Conservative 165; Mismatches 253; Indels 0; Gaps 0;

## FEATURES

## Source

1. 1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BACR48P19"  
/clone\_lib="RPCI-98"  
/note="end : TERT"

BASE COUNT 469 a 6 c 69 g 151 t 406 others

## ORIGIN

## JOURNAL

## COMMENT

Unpublished (2001)  
Other\_GSSs: AG-ND-132P6.TF  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjoftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
Seg primer: MJ3 Rev  
Class: BAC ends.

## FEATURES

## Source

1. 187  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-132P6"  
/clone\_lib="ND-TM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 53 a 42 c 30 g 62 t

## ORIGIN

Query Match 2.9%; Score 54.4; DB 17; Length 187;  
Best Local Similarity 56.0%; Pred. No. 0.0025;  
Matches 103; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

## FEATURES

## Source

1. 187  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-132P6"  
/clone\_lib="ND-TM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 53 a 42 c 30 g 62 t

## ORIGIN

Query Match 2.9%; Score 54.4; DB 17; Length 187;  
Best Local Similarity 56.0%; Pred. No. 0.0025;  
Matches 103; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

## FEATURES

## Source

1. 187  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-132P6"  
/clone\_lib="ND-TM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 53 a 42 c 30 g 62 t

## ORIGIN

Query Match 2.9%; Score 54.4; DB 17; Length 187;  
Best Local Similarity 56.0%; Pred. No. 0.0025;  
Matches 103; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

## FEATURES

## Source

1. 187  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-132P6"  
/clone\_lib="ND-TM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 53 a 42 c 30 g 62 t

## ORIGIN

INSERM U 167  
Institut Pasteur de Lille  
1 rue du Professeur A. Calmette, 59019-Lille, France  
Tel: (33) (0)3 20877783  
Fax: (33) (0)3 20877888  
Email: Raymond.Pierce@pasteur-lille.fr  
CNS sequencing ID=DG0AA022CA04BFI  
Plate: 022 row: B column: 07  
Seq primer: M13 reverse primer  
Class: BAC ends  
High quality sequence stop: 602.  
Location/Qualifiers  
1..602  
/organism="Schistosoma mansoni"  
/strain="Puerto-Rican"  
/db\_xref="taxon:6183"  
/clone="022B07"  
/clone\_lib="SmbAC1"  
/sex="mixed"  
/dev\_stages="cercariae"  
/lab\_host="Biomphalaria glabrata"  
/note="Vector: pBelOBAC 11; Site 1: Hind III; Partially  
Hind III digested and size-selected S. mansoni cercarial  
DNA was ligated into Hind III digested pBelOBAC 11 vector  
and used to transform E. coli DH10B. The complete library  
contains 23808 clones from 4 independent  
sizing-ligation-transformations. Average insert size  
ranges from 70-127 kb and genome coverage is 7.9-fold."

BASE COUNT 194 a 88 c 121 g 158 t 41 others  
ORIGIN

Query Match 2.8%; Score 51.8; DB 17; Length 602;  
Best Local Similarity 42.7%; Pred. No. 0.018;  
Matches 213; Conservative 5; Mismatches 274; Indels 7; Gaps 1

Qy 255 CTGATCTTGGGAAAAACAATGTACACAAGGCATTATCTCAGCATGCTGTGGCTTGGCAA 314  
Db 10 CTTCCTTAGGACAAAAGCAATGTATGCAAGTATTTTATCTTCATTTAGTGGATCTCTCT 69  
Qy 315 TGCCTATGTTTTGTAGAGCGGTATATTATAGATTTGGAGGCGTCATCGCTTCGGGAGCTG 374  
Db 70 TWTATGTGTGTGGATGGTGTCTATTACGGTAGAGCGGTTGGTATGCAAAATCGGCGTT 129  
Qy 375 TTCTTCTGATCTTTTGTCTATCTGGCGAGCTCTACAGTATTTGGATCGGCCCTACCT 434  
Db 130 TATTATTAACAGTATTTATTTCTATTTTGGTGTAAATGGCAAGTTTTTGGTTGTATTAAAT 189  
Qy 435 TGTCAGGACTCGCTGGGATTTGTTCTTGTGTATGGGATGGCGTAGATGCAAAATGTTCTTG 494  
Db 190 TGCAGGTATTGCAGGTATCGSATWAACATTAGGTACTGCAAGTTGATGCCACATCATTA 249  
Qy 495 TATTGGAAGATCCGAGAGGAATTTTTATGCTCTCAAGTCTTTAAAAAATCTGTAGAAA 554  
Db 250 AAAAGAAAGAGCAAAAGAGAAATWACGTGAAGGTAAACATTTAGCAGAAGNAGTTGACA 309  
Qy 555 AAGGATAT-----ACCAAGCGCTTTTGAGCGCATTTTTTGGATCTTAAGTACAGTACAGTA 607  
Db 310 TTTCAATAGTTGGAAGGAGCGCANGCGTAAACATAGTAGAAGCAACGTTACGCATATC 369  
Qy 608 TTGGCCCTCAGCACTCTCTTTTCTCTAGATACAGGGCCTATTAAGAGGTTTGTCTTTTGACA 667  
Db 370 TTAACGGNATCANCITTAANNNGCNTNGAACCGGACCCGACCCAGGANTTTCGNNACA 429  
Qy 668 TTGATTTTAGGAATTTTCTCTTCAATGTTTAGGCTCTTTTTCATGACTAAATTTTCTTC 727  
Db 430 ANANCAAMAGGNAANAAAAAANNNANANANANANANANANANANANANANANANANANANAN 489  
Qy 728 ATGCTGTGGATGAATAAGA 746  
Db 490 GACAAAGAGATGCGGAAGA 508

## RESULT 4

CNS0706R	CNS0706R	602 bp	DNA	linear	GSS 10-OCT-2001
LOCUS	T3 end of clone 022CA04 of library SmbAC1 from strain Puerto-Rican				
DEFINITION	of Schistosoma mansoni, genomic survey sequence.				
ACCESSION	AL619845				
VERSION	AL619845.1	GI:16033987			
KEYWORDS	GSS.				
SOURCE	Schistosoma mansoni.				
ORGANISM	Schistosoma mansoni				
REFERENCE	1 (bases 1 to 602)				
AUTHORS	Strigedida; Metazoa; Platyhelminthes; Trematoda; Digenea;				
TITLE	1 (bases 1 to 602)				
JOURNAL	Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,				
MEDLINE	Williams,D.L., Johnston,D., Loverde,P.T. and Le Paslier,D.				
PUBMED	Construction and characterization of a Schistosoma mansoni				
REFERENCE	bacterial artificial chromosome library				
AUTHORS	Genomics 65 (2), 87-94 (2000)				
TITLE	2 (bases 1 to 602)				
JOURNAL	Genoscope.				
MEDLINE	Submitted (05-OCT-2001) Genoscope - Centre National de Sequences :				
PUBMED	BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr				
REFERENCE	- Web : www.genoscope.cns.fr)				
AUTHORS	Partially Hind III digested and size-selected S. mansoni cercarial				
TITLE	DNA was ligated into Hind III digested pBelosAC 11 vector and used				
JOURNAL	to transform E. coli DH10B. The complete library contains 23808				
COMMENT	clones from 4 independent sizing-ligation-transformations. Average				
FEATURES	insert size ranges from 70-127 kb and genome coverage is 7.9-fold.				
source	Location/Qualifiers				
BASE COUNT	194 a 88 c 121 g 150 t 41 others				
ORIGIN	1. .602				
Query Match	/organism="Schistosoma mansoni"				
Best Local Similarity	/strain="Puerto-Rican"				
Matches 213; Conservative	/db xref="taxon:6183"				
	/clone="022CA04"				
	/clone_lib="SmbAC1"				
	/notes="end : T3"				
Qy 255	CTGATCTGGGAAAAAACAATGTATACAAAGCATATCTTCAGCATGCTGTGGCTTGGCAA 314				
Db 10	CTTCTTTAGGACAAAAAGCAATTGATGAGGTATTTTATCTTCATTAGTTGGATTCTTCT 69				
Qy 315	TGCTTATGTTTGATGAGCGTATATTATAGATTGGAGCGTCATCGCTTCGGGAGCTG 374				
Db 70	TWVTATGTGTGGATGGTGTCTTATTACGGTAGAGCGGGTGGTATGCCAATGCGGCTT 129				
Qy 375	TTCTTCTGAATCTTTTGCTTATCTGGGCAGCTCTACAGTATTTGGATGCGGCACCTCACCT 434				
Db 130	TATTATTAAAGTATTATCTTATTGGTGAATGGCAAGTTTGGTTTGTATTAAACAT 189				
Qy 435	TGTCAGGACTCGCTGGGATGTTCTTGCTATGGGGATGGCCGTAGATGCAAAATGTTCTTG 494				
Db 190	TGCCAGGTATTGCAAGTATCGSATWAAACATTAGGTACTGCAAGTTGATGCCAACATCATTA 249				
Qy 495	TATTTCGAAGNATCCGAGAGAAATTTTATTGTTCTCAAAGTCTTAAAAAATCTGTAGAA 554				
Db 250	AAAHGAAGAGCAAAAGAAAGATWACGTGAAGGTAAACACTTTAGCAGAGNAGTTGACA 309				
Qy 555	AAGGATAT-----ACCAAGGCTTTTGGAGCCATTTTGTATTCTTAACCTTGACTACAGTA 607				
Db 310	TTTCAAATGGTGAAGAGGAGCCANGCGTAACAATAGTAGAAGCAACAGTTACGCATATC 369				
Qy 608	TGCGCTCAGACATTTCTTTCTTCTTAGATACAGGGCCCTATAAAGGGTTGCTTTTGACA 667				
Db 370	TTAACGGGNATCANCTTAAANNNGNNTNGAAACCCGGAACCCANCCAGGANTTTCGNNACA 429				

37  
up

QY	668	TTGATTATTAGGAAATTTCTCTTCAAGCTTTACGGCTTTTCAATGACTAAATTTTCTTC	727
Db	430	ANANCAAAAGGANAANAAAAACNNCANNANANACNNCAACCACGCAAGAACNANCMANC	489
QY	728	ATGCTGTGATGAATTAAGA	746
Db	430	GACCAAGACATGGCGAAGA	508
RESULT 5			
	BH397261/	BH397261	731 bp DNA linear GSS 11-DEC-2001
LOCUS	AG-ND-147L1.TR	ND-TAM	Anopheles gambiae genomic clone AG-ND-147L1,
DEFINITION	DNA sequence.		
ACCESSION	BH397261		
VERSION	BH397261.1	GI:7343477	
KEYWORDS	GSS.		
SOURCE	African malaria mosquito.		
ORGANISM	Anopheles gambiae		
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;		
	Anopheles.		
REFERENCE	1 (bases 1 to 731)		
AUTHORS	Shelley,J., Maleh,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J		
TITLE	Direct Submission of BAC-end sequences from Anopheles gambiae		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Brendan J Loftus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: b.loftus@tigr.org This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PBST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest. Seq primer: M13 Rev Clas: BAC ends. Location/Qualifiers 1..731 /organism="Anopheles gambiae" /strain="PBST" /db_xref="taxon:7165" /clone="AG-ND-147L1" /clone_1lb="ND-TAM" /note="Vector: pECBAC1; Site_1: HindIII"		
BASE COUNT	213 a	146 c	127 g 245 t
ORIGIN			
Query Match	2.7%;	Score 50.8;	DB 17; Length 731;
Best Local Similarity	52.7%;	Pred. No. 0.035;	
Matches 155;	Conservative 0;	Mismatches 137;	Indels 2; Gaps 2
QY	438	CAGGACTCGCTGGGATTTCTTGTGATAGGGGATGGCCGTAGATGCAANGTCTTGAT	497
Db	729	CAGGATTCGGGGTATCGATTAATCATGCGCCAGGACAGTAGACGAACGTATATTT	670
QY	498	TGGAAGAATCCGAGAGGAATTTTATTTCTCAAGTCTTAAAAATCTGTAGAAAAG	557
Db	669	ATGAAGAAGCAAGAGAAATTTTTCGACAGAAAAGGAATCCTGAAGATACACAGCAG	610
QY	558	GATATACCAAGGCTTTTGAGCCATTTTGTATTTACTGACTACAGATATGGCCTCAG	617
Db	609	GTTTCAAGCAGCCTTATATGCGATTATTAACGACACGCACACAACATTAATACACAG	550
QY	618	CACAT-TCCTTCTCTCTAATACAGGCGCTATTAAGGGTTTGCTTG-ACATTAATTTT	675

Db	549	TGTTGATATATATATTTTGGTAGACAGACCAATCCAAAGATTTGGCGAAACCTTAATAT	490
Qy	676	AGGAATTTCTCTTCATGTTTACGGCTCTTTTCACTACTAATTTTCTTCAT	729
Db	489	CGGATATTTGATGACGTTCTTACTTCTGTATATATTTGCGAGATTAATGATCTT	436
RESULT 6			
CNS00FOO			
LOCUS			
DEFINITION	CNS00FOO	1101 bp	DNA
ACCESSION			linear
VERSION			GS5 03-JUN-1999
KEYWORDS			
SOURCE	AL070854.1	GI:4950896	
ORGANISM	GS5.		
REFERENCE	Drosophila melanogaster.		
AUTHORS	Drosophila melanogaster		
TITLE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Prexygota;		
JOURNAL	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
	Ephydroidea; Drosophilidae; Drosophila.		
	1 (bases 1 to 1101)		
	Genoscope.		
	Direct Submission		
	Submitted (02-JUN-1999) Genoscope (E-mail: seque@genoscope.cns.fr		
	Bp 191 91006 EXRX cedex - FRANCE (E-mail: seque@genoscope.cns.fr		
COMMENT	- Web : www.genoscope.cns.fr		
	Determination of this BAC-end sequence was carried out as part of a		
	collaboration with the Berkeley Drosophila Genome Project (BDGP).		
	The BDGP is constructing a physical map of the Drosophila		
	melanogaster genome using these BACs. For further information		
	please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila		
	melanogaster BAC library was prepared by Kazutoyo Osoegawa and		
	Aaron Mammoeer in Pieter de Jong's laboratory in the Department of		
	Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,		
	NY. The library is named RPCI-98 and was constructed by partial		
	EcoRI digestion of Drosophila DNA provided by the BDGP from the		
	isogenic strain y2; cn bw sp, the same strain used for the BDGP's		
	P1 and EST libraries. A more detailed description of the library		
	and how to order individual BAC clones, the entire library, or		
	filters for hybridization from the BACPAC Resource Center can be		
	found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .		
FEATURES			
source	1..1101		
	/location/Qualifiers		
	/organism="Drosophila melanogaster"		
	/db_xref="taxon:7227"		
	/clone="BACR31F03"		
	/clone_1b="RPCI-98"		
	/note="end : T7"		
BASE COUNT	167 a 84 c 86 g 155 t		609 others
ORIGIN			
Query Match	2.6%; Score 49.2; DB 17; Length 1101;		
Best Local Similarity	14.2%; Pred. No. 0.11;		
Matches	61; Conservative 186; Mismatches 182; Indels 0; Gaps 0;		
Qy	491	CTTGATTCGAAAGATCCGAGAGGATTTTATTTGTTCTCAAGTCTTAAATAATCTGTA	550
Db	656	MTMTTNGKVKMMNNMMNNHCAVATKATATGTMGFMKGGWKRTAAMAAAMGMMHRYVM	715
Qy	551	GAAAGAGATATACCAAGCGCTTTTGAGACCATTTTTGATTTCTAATCTGACATACATAT	610
Db	716	GRGAAGCKMKMKVVRMMKMTVRWVGHHGTTKHHBKHHBTKYKTBTTTCCGCGSKKKG	775
Qy	611	GGCTCAGCAGCTCTTTCTTCTAGATACAGGAGGCTTATTAAGAGGTTGCTTGACATTTG	670
Db	776	TKMMMMHETHTVTHGIVTMMTTNGMKMGMMGMMGMMGMMGMMGMMGMMGMMGMMGMMG	835
Qy	671	ATTTTAGGATTTTCTCTCAATGTTTACGGCTCTTTTCAATGACTAAATTTTCTTCAT	730
Db	836	KNGTKNNMMGTMTDMMTKTTTKTKTKTATATWTGTMKKKKDKKKKKTKAKKKM	895
Qy	731	CTGTGATGAATTAAGACCAACATPACACAGTTGACATATGATGAATTAAGTTCTGTGGGATA	790

```

Db 896 KKKAKKAGGAGGAKDAKATAAAWAAKAKAKAADAADAAAADAAKADKGRWDGA 955
Qy 791 AAGCATGATTTCTTGAGAGGATGCAAAAACCTTTGGGCTGTTTCTGGAAGTGTTTCTT 850
Db 956 KAKRTKDAKARAKADAKADRAKDKKKKKKKKKKKKKKKKKKKKKKKKKKKKKGA 1015
Qy 851 TTAGGTCGCTGCTCGGTTTGGAGCGTGGAAATCCGTTTGGGAATGATTTAAA 910
Db 1016 KAKAKKATKKKKDKAKDKAKDKAKDKKKKKKKKKKKKKKKKKKKKKKKKKKK 1075
Qy 911 GGAGGGTAT 919
Db 1076 KKKKTATATK 1084

RESULT 7
CNS0039G      1101 bp      DNA      linear      GSS 03-JUN-1999
LOCUS      Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION      BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
ACCESSION      AL063921
VERSION      AL063921.1 GI:4941778
KEYWORDS      GSS.
SOURCE      Drosophila melanogaster.
ORGANISM      Drosophila melanogaster.
AUTHORS      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
TITLE      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
JOURNAL      Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 1101)
AUTHORS      Genoscope.
TITLE      Direct Submission
JOURNAL      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
COMMENT      BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES             Location/Qualifiers
     source           1..1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      201 a      64 c      131 g      202 t      503 others
ORIGIN

Query Match      2.6%; Score 48; DB 17; Length 1101;
Best Local Similarity 21.0%; Pred. No. 0.22;
Matches 98; Conservative 161; Mismatches 207; Indels 0; Gaps 0;

Qy 1397 TTCCTTTTGAAGAAATCAATAGATTGCGAGCATTGGTCTTTTAATGACTGATTG 1456
Db 403 WWWWWTTTTTTTAAWAAAAAATAATTTAAWAAAAAATAATTTAAWAAAAAATAAT 462
Qy 1457 GGGTATTCATTAACAATCTTTGATCATTTTGTATCGTATTTCGGAAGATCGCAACG 1516
Db 463 WTTTAAWAAAAAATAATTTTAAWAAAAAATAATTTTAAWAAAAAATAATTTTAAW 522

```

```

Qy 1517 AACCTGTTTACCCCTATGCATGTTTGTAGTTAATGATGCCCTTCAAAAGACGCTT 1576
Db 523 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 582
Qy 1577 ACGGTAATGACAAACAGCTACAACTCTATCAGTTTGTATGCTTTTGTATAGGCGCG 1636
Db 583 YTYWYTWTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHY 642
Qy 1637 TCCTCTGCTTTTAAATTTTGCATTTTATGACCATAGGGGATTTCTTCTAGGAAC 1696
Db 643 YTCMYVYHYHMHHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 702
Qy 1697 TCCTCTTATATGACCACTCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1756
Db 703 YXXHHYTAHHTHTHHYAHYHYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMY 762
Qy 1757 TAAGTACCGTTAAACTTAATCTAAGCTGTAGCAATATAAAATCTCTTCTGGGACT 1816
Db 763 AMWWHHHAHYAAAAAATAATTHHHYHHTTHYHMYHYMYHYMYHYMYHYMYHYMYH 822
Qy 1817 TCCCAAGGCGCCCTGTGTTATTAATTTATGACAAATTCAGATAAT 1862
Db 823 YTCWYTHHWWMTWTHYHHTHWHHTTTHHAAWMTHTWCWWWHAT 868

RESULT 8
CNS0100X      1101 bp      DNA      linear      GSS 26-JUL-1999
LOCUS      Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION      BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit
ACCESSION      AL098379
VERSION      AL098379.1 GI:5609990
KEYWORDS      GSS.
SOURCE      Drosophila melanogaster.
ORGANISM      Drosophila melanogaster.
AUTHORS      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
TITLE      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
JOURNAL      Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 1101)
AUTHORS      Genoscope.
TITLE      Direct Submission
JOURNAL      Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
COMMENT      BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.

FEATURES             Location/Qualifiers
     source           1..1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN03G04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBelOBAC11"
                     /note="end : SP6"
BASE COUNT      195 a      108 c      131 g      161 t      506 others
ORIGIN

Query Match      2.5%; Score 47.2; DB 17; Length 1101;
Best Local Similarity 20.9%; Pred. No. 0.36;
Matches 82; Conservative 143; Mismatches 168; Indels 0; Gaps 0;

Qy 519 TTTTATTGCTCAAGCTTAAAAAATCTGAGAAAAGGATATACCAAGCTTTTGAG 578
Db 325 TTTTNTTNTTBDAGGWATARWAWDCRYATATAAAADAACCTTTTGTGTDTKTTT 384
Qy 579 CCATTTTGTATTCTAACTTGACTACAGTATTGGCCTCAGCACTTCTTTTCTCTAGATA 638

```

[illegible]

RESULT 9	
CNS000D1	CNS000D1
LOCUS	1101 bp DNA linear GSS 03-JUN-1995
DEFINITION	Drosophila melanogaster genome survey sequence TERT end of BAC # BAR01J16 of RPT-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL065414
VERSION	AL065414.1
KEYWORDS	GI:4938827
SOURCE	GSS.
ORGANISM	Drosophila melanogaster.
	Drosophila melanogaster.
	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
	Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1101)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage

COMMENT

Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kautzo Oseguera and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES	source	location/Qualifiers
	1..1101	
		/organism="Drosophila melanogaster"
		/db_xref="taxon:7227"
		/clone="BACR01j16"
		/clone_11b="RBCI-98"
		/note="end : TERT3"
BASE COUNT	280 a	104 c 123 g 211 t 383 others
ORIGIN		

Query Match	2.5%	Score 46.8;	DB 17;	Length 1101;
Best Local Similarity	15.7%;	Pred. No. 0.45;		
Matches 67;	Conservative 183;	Mismatches 178;	Indels 0;	Gaps 0;

```

Oy 484 AAATGTCCTTGATTCGAAAGAAATCCGAGAGAAATTTTATTTGTCGCAAGCTTAAAAA 543
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 625 AAMHAGTMMTKDPTKKKKKKKKKKMGAKARAAATDPAWTATTTWKTKAKDAGCAAKKADR 684
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 544 ATCTGTAGAAAAAGATATTAACCAAGCTTTTGAGCCATTTTGAATCTAACTTGACTAC 603
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 685 TKAMGAGAGAGARKATAPAKAKKMKDDPTKKKKTXTATTKKKTITTKRAATDKMKWR 744
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 684 AGTATGGCCGACGACCTTTTCTTCTGATACAGGGCCTATTAAGGTTTGCTTT 663
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 745 KKKATAGAMDDADARGGRGEMDRRTTGAAGDXTGKKKKTDDKDKGGGDDRGDG 804
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 664 GACATTTGATTTTAGCAATTTCTCTTCATGTTTACGGCTCTTTTCACATAAATTTT 723
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 805 GAGGGDRDRDARARABMDTKDTRGAKDADKKKAGRKRMWTRTKWTRDPAWMAAM 864
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 724 CTTATGCTGTGATGAATAAGACCAACATACACAGTTGACTATGATGATTAAGTTGCT 783
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 865 RRRMRDRDKRRGRDRRRMRWAGDKKKMKDDDDTPAKDDWTXKDDPWTAAGATGDDDDGA 924
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 784 GGGGATTAAGCATGATTTCTTGAGAGATGCAAAAACTTTGGCGCTTTTCGAGAGTGT 843
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 925 KDGPMKGTDMRTWTKRTKGMGRKDDKGRKTAAGRAGDDKDKKGDKORTDPAKATDGD 984
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 844 TTTTCTTTTAAAGTCGTTGCTCTCGGCTTGAGACCTCGAAATTCGTTTGGGAATGA 903
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 985 TWRDTPDAKRRRRRRRRRAGRGDRKDKKKRMPARWRKKKKKTKDKDDDDGKRWGPA 1044
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy 904 TTTTAAAG 911
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1045 KAKADPAK 1052
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 10  
B353037/c  
LOCUS  
DEFINITION B353037 685 bp mRNA linear EST 06-MAR-2002  
B353037 Dictyostelium discoideum cDNA library, Af Dictyostelium  
discoideum cDNA clone dda48c21 3', mRNA sequence.  
B353037  
B353037 B353037.1 GI:19223549  
EST.  
Dictyostelium discoideum.  
Dictyostelium discoideum.  
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
1 (bases 1 to 685)  
Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.  
Full length cDNA of Dictyostelium discoideum at the aggregation  
stage  
Unpublished (2002)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tehin@genes.nig.ac.jp.

```

FEATURES
source
    Location/Qualifiers
    1..685
    /organism="Dicyostelium discoideum"
    /strain="AX4"
    /db_xref="taxon:44689"
    /clone="qda48c21"
    /clone_11b="Dicyostelium discoideum CDNA library, AF"
    /dev_mat_A="
    /dev_mat_A"
    /dev_stage="Aggregation stage"
BASE COUNT
    199 a
    122 c
    116 g
    247 t
    1 others
ORIGIN

```

Query	836	GGAAGTCTTTTCCTTTAGGTCGCTGCTCCGGGTTTGGAGCCCGGAATCCGTTTG	895
Query Match	2.5%	Score 46.6; DB 13; Length 685;	
Best Local Similarity	50.0%;	Pred. No. 0.43;	
Matches 115; Conservative	0;	Mismatches 115; Indels 0; Gaps 0	





[illegible][illegible]



Query Match 2.3%; Score 43.8; DB 9; Length 1125;  
 Best Local Similarity 43.0%; Pred. No. 2.8;  
 Matches 108; Conservative 11; Mismatches 132; Indels 0; Gaps 0;

QY 484 AAAGTCTTGTATTCGAAAGATCCGAGAGAAATTTTATGTCCTCAAGCTTAATAA 543  
 DB 891 AAAATTTTAAATATWTAATAAAATWMAAAATATTTTATTAATAAAATTAATAA 832  
 QY 544 ATCTGAGAAAAGATATACCAAGCTTTGGAGCCATTTTGAATCTGAAGTAC 603  
 DB 831 TWTATATAAATAATWTAATAAAATTTTATTAATAAAATTTTATTTTNNATTTTN 772  
 QY 604 AGATTTGGCCCTCAGACTCTTTCTTCTCTAGATACAGGCTTAAGGTTTCTT 663  
 DB 771 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 712  
 QY 664 GACATGATTTTAGGAATTTCTCTCAATGTTTACGGCTTTTCATGACATAATTTT 723  
 DB 711 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 652  
 QY 724 CTTCATGCTGT 734  
 DB 651 TTTTATTTTATTTT 641

RESULT 17 417 bp mRNA linear EST 22-MAY-2002  
 BQ399330  
 LOCUS NISC mp03a03.x1 NICHD XGC Spt Xenopus laevis cDNA clone  
 DEFINITION IMAGE:4965412.3, mRNA sequence.  
 ACCESSION BQ399330  
 VERSION BQ399330.1 GI:21087017  
 KEYWORDS EST.  
 SOURCE African clawed frog.  
 ORGANISM Xenopus laevis  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 Xenopodidae; Xenopus.  
 1 (bases 1 to 417)  
 REFERENCE NIH-XGC http://image.llnl.gov/image/html/xenopuslib info.shtml.  
 AUTHORS National Institute of Child Health and Human Development, National  
 Cancer Institute, Xenopus Gene Collection  
 TITLE Unpublished (2002)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgaabs-remail.nih.gov  
 cDNA Library Preparation:  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC)  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 info@image.llnl.gov  
 plate: LLAM10941 row: B column: 5  
 Seq primer: -21M13 forward primer (ABI).  
 Location/Qualifiers  
 1. 417

FEATURES  
 source  
 Location/Qualifiers  
 1. 417  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 /clone="IMAGE:4965412"  
 /clone\_lib="NICHD XGC Spt"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: spleen; Vector: PCMV-SORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 2.4 kb. Constructed by Life  
 Technologies."

BASE COUNT 174 a 56 c 2 g 185 t

ORIGIN

Query Match 2.3%; Score 43.6; DB 14; Length 417;  
 Best Local Similarity 49.1%; Pred. No. 2.2;  
 Matches 115; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 1608 TTTTGTAAATGCTTTTGTATAGGGCTCTCTGCTTAATTTGCAATTATATGA 1667  
 DB 31 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 90  
 QY 1668 CCATAGGATTTCTTTCGAACTTATCGTCTTATATGACACACTCTGTGTGT 1727  
 DB 91 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 150  
 QY 1728 TTATGTCCTGTAAGAAATCGCTCAAAATAGTACCGTTAACTTAATCTAACGTTAG 1787  
 DB 151 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 210  
 QY 1788 CAATATAAATCTCCTTGGACTTATGTCCTCAAGGCCCTGTGTATTA 1841  
 DB 211 AAAAAAAAAATATAATTTTAACTTTACCCCCCCCCCTTCTTAATAA 264

RESULT 18 500 bp mRNA linear EST 27-JAN-2001  
 AU087215  
 LOCUS AU087215 Sugano Malaria cDNA library Plasmodium falciparum 3D7 cDNA  
 DEFINITION clone XPN4019, mRNA sequence.  
 ACCESSION AU087215  
 VERSION AU087215.1 GI:12389356  
 KEYWORDS EST.  
 SOURCE Plasmodium falciparum 3D7.  
 ORGANISM Plasmodium falciparum 3D7  
 Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 REFERENCE 1 (bases 1 to 500)  
 AUTHORS Watanabe, J., Sasaki, M., Suzuki, Y. and Sugano, S.  
 TITLE FUL-malaria: a database for a full-length enriched cDNA library  
 from human malaria parasite, Plasmodium falciparum  
 Nucleic Acids Res. 29 (1), 70-71 (2001)  
 JOURNAL 20574754  
 MEDLINE  
 COMMENT Contact: Junichi Watanabe  
 Institute of Medical Science  
 The University of Tokyo, Department of Parasitology  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Tel: 81-3-5449-5378  
 Fax: 81-3-5449-5410  
 Email: jwatanabe@med.s.u-tokyo.ac.jp  
 Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano  
 S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).  
 Location/Qualifiers  
 1. 500  
 /organism="Plasmodium falciparum 3D7"  
 /strain="3D7"  
 /db\_xref="taxon:36329"  
 /clone="XPN4019"  
 /clone\_lib="Sugano Malaria cDNA library"  
 /dev\_stage="erythrocytic stage"  
 27 c 23 g 274 t 4 others

BASE COUNT 172 a 27 c 23 g 274 t

Query Match 2.3%; Score 43.6; DB 9; Length 500;  
 Best Local Similarity 47.7%; Pred. No. 2.3;  
 Matches 124; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 484 AAAGTCTTGTATTCGAAAGATCCGAGAGAAATTTTATGTCCTCAAGCTTAATAA 543  
 DB 208 AAAATTTTAAATATWTAATAAAATWMAAAATATTTTATTAATAAAATTAATAA 267  
 QY 544 ATCTGAGAAAAGATATACCAAGCTTTGGAGCCATTTTGAATCTGAAGTAC 603  
 DB 268 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 327  
 QY 604 AGATTTGGCCCTCAGACTCTTTCTTCTCTAGATACAGGCTTAAGGTTTCTT 663  
 DB 328 ATGTTTGGATTTATATCTTTCTTTCTTAATATACGAATCGCTTGTGTTTCTT 387  
 QY 664 GACATGATTTTAGGAATTTTCTCTCAATGTTTACGGCTTTTCATGACATAATTTT 723



QY 1405 GAAGAAATTCAGATTTGCAAGCATTGGTCTTATGATCTGATTTGGGATATTC 1464  
 DB 726 TKASHSTTBSCAACSAMTTTTTMMKMMCAACAANKKKMAAAAACAAAKKKKMTAM 785  
 QY 1465 ATTAACAAATCTTGAATTTTGTATTCGTGATGAGATGCCAAGCAACTGTT 1524  
 DB 786 AATWCAATTAADTKTTTATTTTAAAKKKRKBDMMAAMCCCCAKTKTKTTTTT 845  
 QY 1525 TACCCC 1530  
 DB 846 TCSCCC 851

RESULT 21  
 CNS00DM2/c  
 LOCUS  
 DEFINITION  
 Drosophila melanogaster genome survey sequence TERT3 end of BAC # BACR27016 of RP11-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION  
 AL072713.1 GI:4948210  
 SOURCE  
 Drosophila melanogaster.  
 ORGANISM  
 Drosophila melanogaster.  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 603)  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 Direct Submission  
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : segr@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org>. The BDGP Drosophila melanogaster BAC library was prepared by Kazuo Osoegawa and Aaron Mammoler in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RP11-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
 source  
 1..803  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone="BACR27016"  
 /clone\_lib="RP11-98"  
 /note="Tend : TERT3"

BASE COUNT  
 248 a 94 c 131 g 158 t 172 others

ORIGIN  
 Query Match 2.3%; Score 43.2; DB 17; Length 803;  
 Best Local Similarity 31.9%; Pred. No. 3.5;  
 Matches 150; Conservative 60; Mismatches 260; Indels 0; Gaps 0;

QY 1212 GGTCAAGCTAAGCAGCAATCTATCGAAGAAATGCTTATCAGCGACCATCGGCTT 1271  
 DB 483 CRTGAGGAAAAAARFGAAAAAATAAAWAAAGBWTDDTTTGTAKTAAACGRRAC 424  
 QY 1272 TAGAGCTTTGGCAATCATCTGCTCATGTGAGTTTGCCCTTGAATGCAATATGCTT 1331  
 DB 423 AATACCCATKAGWTGTTTTTGTGKTTTWTWTKTKKCCACCNCCNNAATGCTGTTT 364  
 QY 1332 TCAATGCCGTATGCGCTTAAATCATGACCTTTGGCTACCTGTGACGCTTTATAG 1391  
 DB 363 TCTCTGACAGTGTGTTTATATGTCATATGCTTAAACATTTTNNNNNNNTCCCTTTT 304  
 QY 1392 CACATTTCTTTGAAGAAATTCATATGATTTGCAAGCATTTGGCTTATATGCTG 1451  
 DB 303 CTTTNNNNCTTTTNNNNNNNCAANWACCTTTTNNNNATCTTNNATCTTNNACA 244  
 QY 1452 TATGGGGATTCATTAACATATCTTGATCATTTTGTATGCTATTCGT 1501  
 DB 243 TATCTTCTCAANTATTAATAAACAACCTTCATCTTTCCTNNATCAT 194

RESULT 22  
 AZ550729  
 LOCUS  
 DEFINITION  
 Entomocba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.  
 AZ550729  
 ACCESSION  
 AZ550729.1 GI:11176030  
 VERSION  
 KEYWORDS  
 GSS.  
 SOURCE  
 Entamoeba histolytica.  
 ORGANISM  
 Entamoeba histolytica.  
 Eukaryota; Entamoebidae; Entamoeba.  
 1 (bases 1 to 863)  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjoftus@tigr.org  
 Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library  
 HM1:IMSS sheared DNA library  
 Unpublished (2000)

FEATURES  
 source  
 1..863  
 /organism="Entamoeba histolytica"  
 /strain="HM1:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: pHOS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Baret, Oxford University Press, 1999)."

BASE COUNT  
 368 a 62 c 87 g 346 t

ORIGIN

```
Query Match      2.3%; Score 43.2; DB 17; Length 863;
Best Local Similarity 47.1%; Pred. No. 3.6;
Matches 132; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 523 ATTGCTCAAGTCTTAAATAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGCCAT 582
DB 334 ATGAATTATCTATTATGATATTAGAGAGAAATGATATCAAGTAATATATAGTATA 393
QY 583 TTTTGATCTTAACCTGACTACAGTATTGGCTCAGCACTCTCTTTTCTTCTAGATACAGG 642
DB 394 TTACAAATATATGATGAATAAATATTATTAAATAAATACATTTTCATCAATGATCATT 453
QY 643 GCCTATTAAAGGTTGCTTTGACATTTAGGATTTTCTCTTCAATGTTTACGGC 702
DB 454 TTGTAGTATGATATTGATTTATTTCAATTTATGTAATATATCAATTTATTTATC 513
QY 703 TCTTTTCATGACTAAATTTTCTTCATGCTGTGATGAATAAGCAACCAATACACAGTT 762
DB 514 TTTTCATCTTCATTAAATATCAATTTAAATAGATATATTAACAATATGAAACAATATA 573
QY 763 GCATATGATGAATAAGTTCTGGGGGATAAAGCATGATTTTC 802
DB 574 AATAATGACAAATGATTAATACTAAATAAGAAATATTTTC 613

RESULT 23
BG350812/c
LOCUS
DEFINITION      09BE12 Mature tuber lambda ZAP Solanum tuberosum cDNA, mRNA
ACCESSION      BG350812
VERSION        BG350812.1 GI:13179554
KEYWORDS
SOURCE
ORGANISM        Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE
1 (bases 1 to 332)
NIELSEN,K.L, Crookshanks,M., Emmersen,J. and Welinder,K.G.
EST-sequencing of mature potato tuber (Var. Kuras)
Unpublished (2000)
Contact: Karen G. Welinder
Institut for bioteknologi
Aalborg Universitet
Sohngaardsholmsvej 49, 9000 Aalborg, Denmark
Tel: +45 96358467
Fax: +45 98141808
Email: kgw@bio.auc.dk
Sequenced from the 5' end.
High quality sequence stop: 332
POLYA=No.
Location/Qualifiers
1. .332
/organism="Solanum tuberosum"
/cultivar="Field grown Kuras"
/db xref="taxon:4113"
/clone.lib="Mature tuber lambda ZAP"
/tissue.type="Tuber"
/notes="Vector: Lambda ZAP"

BASE COUNT      69 a 60 c 104 g 99 t
ORIGIN

Query Match      2.3%; Score 42.8; DB 12; Length 332;
Best Local Similarity 56.3%; Pred. No. 3.3;
Matches 80; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1457 GGGTATTCAATTAACAATCTTTCATCATTTTTCATCGTATTCGTGAAGATCGCAAGCG 1516
DB 146 GAGCAGTATTCAACACGGGCTTCATCCCTTTAGCTCCAAAAGCACAAGATCCACAGAT 87
QY 1517 AACCTGTTACCCCTATGCATGTTTTCATGATGCGCTTCAAAGACGTTTCAGCGCG 1576
```

```
DB 86 AGCTTCTTCAACCCCTTACAAATCTTCGCAAAAGATTCCTCAGTATCAGTAAGCTCC 27
QY 1577 ACGTAAATGCAACAGCTACAA 1598
DB 26 CGGSCCATGACCGGAGCTTCAA 5

RESULT 24
CNS0155H
LOCUS
DEFINITION      1001 bp DNA linear GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN13C23 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION      AL105023
VERSION        AL105023.1 GI:5617037
KEYWORDS
SOURCE
ORGANISM        Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Phyloidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1001)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaut at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11
Location/Qualifiers
1. .1001
/organism="Drosophila melanogaster"
/db xref="taxon:7227"
/clone.lib="DrosBAC"
/plasmid="pBelobAC11"
/notes="end : SP6"

BASE COUNT      266 a 219 c 134 g 150 t 232 others
ORIGIN

Query Match      2.3%; Score 42.6; DB 17; Length 1001;
Best Local Similarity 29.7%; Pred. No. 5.5;
Matches 126; Conservative 85; Mismatches 213; Indels 0; Gaps 0;

QY 1348 TTAAATTCATGACCTTTTGGCTACCTGTCAGTCTTGTATAGCACATTTCTTTTGA 1407
DB 911 WWWAAWTTTAAATTTATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTAA 852
QY 1408 GAAATTCAAATAGATTGCAAGCCATTGGTGCTTTTAAATGACTGATTTGGGGTATTCATT 1467
DB 851 AAAATTTAAATTTTWTWTAAWTTTWTAAATTTTWTAAATTTTWTAAATTTT 792
QY 1468 AACAATACTTTGATCATTTTGTATGCTATTCGTGAAGATCGCCAGGACCTGTTTAC 1527
DB 791 TAWMTTTTATTTTAAWAAHTTTTATTTTWWMTTAAWTTAAATTTT 732
QY 1528 CCCTATGCATGTTTTCAGTAAATGATGCCCTTCAAAGACGTTTCAGCCGCAAGTAAAGAC 1587
DB 731 HYTYTWTATTTATTTATTTATTTTTHYTYTWTWTWTWTWTWTWTWTWTWTWT 672
QY 1588 AACAGCTACAACTCTATCAGTTTGTAAATGCTTTTGTATAGGCGCTCCCTCTGCTT 1647
DB 671 MCYTTTWTDBNTTTTGTGTTTTHWTHWTWTWTWTWTWTWTWTWTWTWTWT 612
QY 1648 TAATTTGCAATTTATATGACCATAGGATTTCTCTAGGAACATTTATCGTCTCTTATAT 1707
DB 611 MAMYTTTSTYCTHVTTTTCYCMTTTTTCTCTNTATTTCTTCTTCTTCTTCTTCTTCTT 552
```



QY	613	CTCAGCACTTCTTTCTTCTCAGTATACAGGGCCATTAAAGGTTTGCTTGACATTAT	672
Db	304	CTTAGGGGTACATTCCTTACAAATAATAAACAAGTTTAAANTGTTTAGACACAAAT	245
QY	673	TTTAGAAATTTCTCTTCATGTTTACGGCTCTTTTACATGACTAAATTTCTTCATGCT	732
Db	244	ATTGGAAATTTGCTATCTTATGCGCACTGCTGTTTAAATTTTAACTTCACTCAAGGTAT	185
QY	733	GTGAGTAAATAGACCCACATACACAGTTGCTATGATGAAATAGTGTGGGATPAA	792
Db	184	GTAAGTACTACCTTTAAAGTTTATTTAGTGTTATTTTATATCACTACTCGGAAATG	125
QY	793	GCATGATTTCTTGAGAGATGC	814
Db	124	GTTTGTCTTCAAGATGCAATAC	103
RESULT 26			
LOCUS	B80295/c	713 bp	DNA
DEFINITION	CIT-HSP-2045D21.TF CIT-HSP Homo sapiens genomic clone 2045D21, DNA sequence.		
ACCESSION	B80295		
VERSION	B80295.1		
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 713) Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.		
TITLE	Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)		
JOURNAL	Unpublished (1998)		
COMMENT	Other GSSs: CIT-HSP-2045D21.TR Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 0208 Email: mdadams@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tcd/humgen/bac_end_search/bac_end_search.html Seq primer: M13-21 Class: BAC ends.		
FEATURES			
source	Location/Qualifiers		
	1..713		
	/organism="Homo sapiens"		
	/db_xref="GDB:7054560"		
	/db_xref="taxon:9606"		
	/clone="2045D21"		
	/clone_1b="CIT-HSP"		
	/sex="Male"		
	/cell_type="Sperm"		
	/note="Vector: pbelBAC11; site_1: HindIII; site_2: HindIII"		
BASE COUNT	248 a 139 c 119 g 207 t		
ORIGIN			
Query Match	2.3%, Score 42.4; DB 17; Length 713;		
Best Local Similarity	53.7%; Pred. No. 5.5;		
Matches	88; Conservative 0; Mismatches 76; Indels 0; Gaps 0		
QY	487	TGTTCTGTATTCGAAGAAATCCAGAGAAATTTTATGTCTCAAGCTTTAAAAATC	546
Db	479	TGTTTCTAGTTTGGCAAGCTAAGATGCTTTTACATCTCTTAAAGTTGAAAAAAT	420
QY	547	TGTAGAAAAAGATATACCAAGCTTTTGAGCCATTTTATTTCTAACTTACACTACACT	606







```

Db 202 TTAAGACACKTGGCAATTAATTTTGGKTKTKTGGWTTATATWTTTGGKTAAGKTAATWTTT 143
Qy 1817 TCC 1819
Db 142 TTC 140

RESULT 31
CNS04E5M          976 bp   DNA   linear   GSS 21-MAY-2000
LOCUS             Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION        103P02 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION         AL288627
VERSION           AL288627.1 GI:8025084
KEYWORDS           GSS; genome survey sequence.
SOURCE            Tetraodon nigroviridis.
ORGANISM          Tetraodon nigroviridis
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
                  Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
                  Tetraodontidae; Tetraodon.
REFERENCE         1 (bases 1 to 976)
AUTHORS           Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
                  Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
                  Saurin,W. and Weissenbach,J.
TITLE             Human gene number estimate provided by genome wide analysis using
                  Tetraodon nigroviridis DNA sequence
JOURNAL           Unpublished
REFERENCE         2 (bases 1 to 976)
AUTHORS           Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
                  Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
                  Weissenbach,J.
TITLE             Characterization and repeat analysis of the compact genome of the
                  freshwater pufferfish Tetraodon nigroviridis
JOURNAL           Unpublished
REFERENCE         3 (bases 1 to 976)
AUTHORS           Genoscope.
TITLE             Direct Submission
JOURNAL           Submitted (12-APR-2000)
COMMENT           This sequence is a single read and was generated as part of a large
                  scale clone-end sequencing project of the Tetraodon nigroviridis
                  genome. For more information, please take a look at
                  http://www.genoscope.cns.fr/Tetraodon.
FEATURES          Location/Qualifiers
source            1..976
                  /organism="Tetraodon nigroviridis"
                  /db_xref="taxon:99883"
                  /clone="103P02"
                  /clone_lib="G"
                  /notes="Genoscope sequence ID : COBG103DH01LP1-end : T7"
BASE COUNT       243 a 151 c 137 g 327 t 118 others
ORIGIN

Query Match      2.2%; Score 41.8; DB 17; Length 976;
Best Local Similarity 38.2%; Pred. No. 8.8;
Matches 109; Conservative 28; Mismatches 148; Indels 0; Gaps 0;

Qy 517 ATTTTATTGTCGAAGCTCTTAAATAATCTGTAGAAAAGCATATACCAAGGCTTTGG 576
Db 678 ATTTATTWTTTAAATNNAWATTAATAAATTTTAWTTWATATAATTAATTAATTA 737
Qy 577 AGCCATTTTGATCTTAACCTGTACAGTATGCGCCTCAGCACTCTCTTTCTTCCTAGA 636
Db 738 WATWTTWAWATTTTWTWAWATTTWTTWATNATWATWATWATWATTTTATTA 797
Qy 637 TACAGGCGCTATTAAAGGGTTTGCTTGACATTTAGGAATTTTCTCTTCAATGTT 696
Db 798 TTAATAWATTAATAAATAAATTTTATTAATTTTATTAATWATWATWATWATTTT 857
Qy 697 TACGCGCTTTTCAGACTAAATTTTCTTCATGCTGTGGATGAATAGACCCCAACATAC 756

```

```

Db 858 TATTATTWTTWTTNTTANAANAATATATATAATWATTTTNTWAAATTTAATTTTAAATTA 917
Qy 757 ACAGTTGCATATGATGAATAGTTTCGTGGGATAAAGCATGATTT 801
Db 918 TAAATATAATATTATTTTATTTTATTAATAAATAATTAATAAATTTT 962

RESULT 32
CNS051P2          1096 bp   DNA   linear   GSS 26-MAY-2000
LOCUS             Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION        024H02 of library A from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION         AL339167
VERSION           AL339167.1 GI:8232925
KEYWORDS           GSS; genome survey sequence.
SOURCE            Tetraodon nigroviridis.
ORGANISM          Tetraodon nigroviridis
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
                  Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
                  Tetraodontidae; Tetraodon.
REFERENCE         1 (bases 1 to 1096)
AUTHORS           Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
                  Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
                  Saurin,W. and Weissenbach,J.
TITLE             Human gene number estimate provided by genome wide analysis using
                  Tetraodon nigroviridis DNA sequence
JOURNAL           Unpublished
REFERENCE         2 (bases 1 to 1096)
AUTHORS           Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
                  Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
                  Weissenbach,J.
TITLE             Characterization and repeat analysis of the compact genome of the
                  freshwater pufferfish Tetraodon nigroviridis
JOURNAL           Unpublished
REFERENCE         3 (bases 1 to 1096)
AUTHORS           Genoscope.
TITLE             Direct Submission
JOURNAL           Submitted (12-APR-2000)
COMMENT           This sequence is a single read and was generated as part of a large
                  scale clone-end sequencing project of the Tetraodon nigroviridis
                  genome. For more information, please take a look at
                  http://www.genoscope.cns.fr/Tetraodon.
FEATURES          Location/Qualifiers
source            1..1096
                  /organism="Tetraodon nigroviridis"
                  /db_xref="taxon:99883"
                  /clone="024H02"
                  /clone_lib="A"
                  /notes="Genoscope sequence ID : COAA024D01C1-end : T7"
BASE COUNT       488 a 127 c 179 g 236 t 66 others
ORIGIN

Query Match      2.2%; Score 41.8; DB 17; Length 1096;
Best Local Similarity 40.3%; Pred. No. 9.1;
Matches 127; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

Qy 1420 AGATTTCGAAGCCATTGCTGCTTTAATGACTGTATTCGGGTATTCATTAACAATACTTT 1479
Db 455 AGAAGTGGCTGCGTNNNGTACTCACTTTCCACCACATTTGTATCTTTCAACAGAAATCAT 396
Qy 1480 GATCATTTTTCATCGTATTCGTGAAGATCCCAAGCAACCTGTTTACCCCTATGCATGT 1539
Db 395 NATWTGCTTTTNTTTTATNCNTNTNCAGCCTAAAGCAACCCANTTTTNCNNGAAACNTTTT 336
Qy 1540 TTTAGTTAATGATGCCCTTCAAAAGACGTTTCAGCGCACGGTAAATGACACAGCTACAAC 1599
Db 335 TTTTACTCCNTNCAACANNNNNNNNNTTTTTTTTNTNTNTNTTTTTTTTNCNNTNTTTTCAAG 276
Qy 1600 TCTATCAGATTTTGTAAATGCTTTGTTTATAGGGCGGCTCCTGCTCTTTAATTTTGCATT 1659
Db 275 TTTNTTTTTTTTTTTTTTTTNTNTNTNTTTTTTTTTTTTTTTTTTTTTNTNTNTNTTTTTTTT 216

```

Oy	1660	TATTTGACCAATAGGAATCTTCCTTGAACTTATCGTCTTTATATTGCACCACTCT	1719
Dd	215	TT	156
Oy	1720	GTTGTGTTTATGCT	1734
Dd	155	TTTTCTCTTTTTTGGT	141
<hr/>			
RESULT 33			
CNS003B4/C	CNS003B4	1101 bp	DNA linear GSS 03-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence TET3 end of BAC #		
DEFINITION	BACR08G08 of RPc1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL064082		
VERSION	AL064082.1	GI:4941838	
KEYWORDS	GSS.		
SOURCE	Drosophila melanogaster.		
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epmydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101)		
REFERENCE	Genoscope. Direct Submision Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr		
AUTHORS	- Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial ECORI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.		
FEATURES	Location/Qualifiers		
source	1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACR08G08" /clone_1ib="RPc1-98" /note="end : TET3"		
BASE COUNT	573 a 140 c 134 g 144 t	110 others	
ORIGIN			
Query Match	2.2%;	Score 41.8;	DB 17; Length 1101;
Best Local Similarity	39.6%;	Pred. No. 9.1;	
Matches	172;	Conservative 18;	Mismatches 244; Indels 0; Gaps 0;
Oy	477	TGATGCAAAATGTTCTTGTAITTCGAAAAGAATCCGAGAGAAATTTTATGTGCTCAAAGTC	536
Dd	1087	TATATAMATMTATATTTTTTTTATTAATAATTNNWTATATAMAATTTTTTANAAAAAATA	1028
Oy	537	TAAAAAAATCTGNAGAAAAGATATACCAGGCTTTTGAGCATTTTGAATCTACT	596
Dd	1027	TATATTTWTAATATATAATATAATATATTWTTTATTTTTTTTTTTTTTTTWTATATAATT	968
Oy	597	TGACTACAGTATTTGGCTCAGCACCTTCTTTCTCCTAGATAACAGGCCATTAAGGCT	656
Dd	967	NMAAMATWTTTWTWTTTWTWTTTMMMAAAATATATTTTTTTTTTTTANNATTAATNTTTTTTTT	908
Oy	657	TTCCTTGACATTTGATTTTNGAATTTTCTCTTCATGTTTAAAGGCTCTTTTCATGACATA	716

QY	717	AAATTTTCTTCAAGCTGCGATGAATGAAGCCCAACATACGACTTGATGTGATGAATA	776
Db	907	TTTTTTTAAATTAATTTTTTTTTTTAAATTTTAAATTTTAAATTTT	848
QY	717	AGTCGCGGAGATTAAGCATGATTTCTTGAGAGATGCAAAAACCTTTGGCTCTTCTG	836
Db	847	TTTTTTTTTTTTTTTTTTTTTTTTTAAATTAATTAATTAATTAATTTT	788
QY	787	TTTTTTTAAATTTT	728
QY	837	GAAGTGTTCCTTTAGGTTGCTGCTCGGCTTTGAGCGCTGCAATTCCTTTTGG	896
Db	727	TTTTTTTTTTTTTTTTTTTAAATTTTAAATTTTAAATTTT	668
QY	897	GAATGATTTTAA	910
Db	667	TATTAATTTTAA	654
RESULT	34		
LOCUS	AZ680882	930 bp	DNA
DEFINITION	ENT6161TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.		
ACCESSION	AZ680882		
VERSION	AZ680882.1	GI:11818028	
KEYWORDS	GSS.		
SOURCE	Entamoeba histolytica.		
ORGANISM	Entamoeba histolytica.		
REFERENCE	1 (bases 1 to 930)		
AUTHORS	Loftus, B., Van Aken, S. and Fraser, C.		
TITLE	Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Brendan J Loftus Department of Eukaryotic Genomics 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: b.loftus@ictr.org Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library Seq primer: M13-Reverse Class: Shotgun High quality sequence start: 20 High quality sequence stop: 750. Location/Qualifiers 1..930 /organism="Entamoeba histolytica" /strain="HMI:IMSS" /db_xref="taxon:5759" /clone_lib="Entamoeba histolytica Sheared DNA" /note=Vector: pHOSt; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broch cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (-2 Kb). The v + i method used for the library construction is described in detail in Smith, H.O., and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."		
BASE COUNT	386 a	88 c	75 g 381 t
ORIGIN			
Query Match	2.2%	Score 41.6;	DB 17; Length 930;
Best Local Similarity	46.8%	Pred. No. 9.7;	
Matches 131;	Conservative 0;	Mismatches 149;	Indels 0; Gaps 0;

```

Qy 523 ATTGCTCAAGTCTTAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCCAT 582
Db 210 ATGGAATTAATCTATATATGATATTTAGAGAGAAATGAATAATCAATATATATAGTATA 269
Qy 583 TTTTGATTCTAACTTGACTACAGTATTCGCCCTCAGCACTCTCTTTCTTCCCTAGATACAGG 642
Db 270 TTACAAATATATGATGAATAAATATTAATTAATATAAATATATATATATATATATATAT 329
Qy 643 GCCTATTAAAGGTTTGGCTTTGACATTTAGCAATTTTGGCAATTTTCTCTCAATGTTTACGGC 702
Db 330 TTGTAGTATTGATATTGATATTCTTCAATTAATGTAATATATATATATATATATATATAT 389
Qy 703 TCTTTTCATGACTAAATTTTCTTCATGCTGTGGATGATAGAACCCCAACATACAGATT 762
Db 390 TTTTCATCTTCATTAATATCAATTTAAATAGAAATATAATTAACAATAATGAACAATATA 449
Qy 763 GCATATGATGAATAGTTTCGTGGGATAAAGCATGATTTTC 802
Db 450 AATAATGCAAAATGATTAATACTAATAAATGAATATTTTC 489

```

```

RESULT 35
AZ541040 931 bp DNA linear GSS 14-NOV-2000
LOCUS ENTDD56TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION AZ541040
VERSION AZ541040.1 GI:11148393
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 931)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HM1:INSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:INSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 89
High quality sequence stop: 909.
Location/Qualifiers
1..931
/organism="Entamoeba histolytica"
/strain="HM1:INSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHOS1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
```

```

FEATURES
source
1..931
/organism="Entamoeba histolytica"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHOS1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
```

BASE COUNT 262 a 174 c 99 g 396 t

ORIGIN

```

Query Match 2.2%; Score 41.6; DB 17; Length 931;
Best Local Similarity 53.3%; Pred. No. 9.7;
Matches 112; Conservative 0; Mismatches 94; Indels 4; Gaps 1;

Qy 1586 ACAACAGCTCAACCTCATCATGTTTGTGTTAATGCTTTGTTTATAGCGGCTCCTCTGTC 1645
Db 127 ACAAGAGACATACTCAATTAATTTTAAAGGCTAATTTTTTTCCTTATTAATTAAT 186
Qy 1646 TTTAATTTTGCATTTATATGACCATAGGATTTCTTCTAGGAATTTTATCGTCTCTCTTAT 1705
Db 187 CTTAAT---CTTTTTCATATACACAAATTAATTTAAATTCCTTTTCTGTTATTAAT 242
Qy 1706 ATTGCACACCTCTGTTGTTTATGTCGCTAAAGAAATCGCTCAAAATAGTACCG 1765
Db 243 TTATAAAAATCTCCCTTTTGTGTTTTTAAACAAATTAATAAATCTAAACAAACAAATACAA 302
Qy 1766 TTAAACTTAATCTAAACGTAGCAATATAA 1795
Db 303 TTAATAAATATCATTTGTTTACCAGAAATGA 332

```

```

RESULT 36
CNS00EPP 1080 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACR29N07 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL069494
VERSION AL069494.1 GI:4949637
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE 1 (bases 1 to 1080)
AUTHORS Direct Submission
TITLE Drosophila melanogaster
JOURNAL Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org/TheBDGP
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoler in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
PI and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
Location/Qualifiers
1..1080
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="BACR29N07"
/notes="end : TET3"
```

BASE COUNT 171 a 105 c 219 g 225 t 360 others

ORIGIN

```

Query Match 2.2%; Score 41.6; DB 17; Length 1080;
Best Local Similarity 21.6%; Pred. No. 10;
Matches 109; Conservative 151; Mismatches 244; Indels 0; Gaps 0;

```

```

Qy 417 TGGATGCCCACTCACCTTGTGAGGACTCGCTGGGATGTTCTTGTCTATGGGATGCCG 476
Db 453 TGCNNNNNNNNNNCCCTBKGGKGGTTCGNGGGGGGGGGGGGGTNGKGGKKGKG 512

```

Qy	477	TAGATGCAAAATGTTCTTGTATTCGAAAGATCCGAGAGATTTTATATGTCGCAAGTC	536
Db	513	TKTKTKKGGKRCCTTWTWTGGGGGGKGTADSAATTGGTTTTTTTTTTTCTTTTDTAGTK	572
Qy	537	TTAAAAAATCTGTAGAAAAGATATACCAAGGCTTTTGGAGCAATTTTGAATCTTA	596
Db	573	WKAAMRKTKWTWTWTKGTTGAGRWMAAGADWTTTDDAAGATGTRKMGKMGKMK	632
Qy	597	TGACATACAGTATTGGCTTAGCACTCTTTCTTCTTCTATACAGGSCCTATTAAGG	656
Db	633	TKKKTTRGTRKWTYKRGKQDTRKTYGKTXTKTKDGRKKGKGBDTAGATADRKTAKG	692
Qy	657	TTGCTTTGACATGATTTTATAGAAATTTCTCTTCATGTTTACGGCTCTTTTCA	716
Db	693	KRRBRAGGKGDGGTTTDXGDGDMWADATDTWTKKDDTTWMDATDRKMMWGMWAD	752
Qy	717	AATTTTCTTCATGCTGTGATGAATTAAGCCCAACATACACAGTTGCATATGATGA	776
Db	753	DDDMWAGRCAGARADDRIADADARGMAAAKDKTTAMATDTTTRKTRDDDTDTGG	812
Qy	777	AGTTGTGGGGATTAAGCATGATTTCTTGTAGAGATGCAAAAACCTTGGGCTGTT	836
Db	813	RKGKMGKRRDGGGAGTRKAKGARKGCTDAAAAMWMDRTTALXKDTKTTTTTATTT	872
Qy	837	GAAATGTTTTCTTTAGATGCGTTGCTCTCGGGTTTGAGCGCTGGAATTCGTTTGG	896
Db	873	GGAGGDTGNWKATADATATAKRWKDKRDKCAARATAMAKAGDWAKADGKRKR	932
Qy	897	GAAATGATTTTAAAGAGGATG 920	
Db	933	GAAATRRKGAADTTRDWTXGMGR 956	

```

RESULT 37
BF279297
LOCUS
DEFINITION GA_Eb0038F20f Gossypium arboreum 7-10 dpa fiber library Gossypium
ACCESSION BF279297
VERSION BF279297
KEYWORDS
SOURCE
ORGANISM
Gossypium arboreum.
Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 552)
Wing,R.A., Fritsch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
D., Wood,T.C., Jesile,A. and Wilkins,T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2003)
JOURNAL
COMMENT
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATGACTCAGCTACGAG
High quality sequence stop: 547.
Location/Qualifiers
1..552
FEATURES
source

```

```

/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone_id="GA_EB003BF20f"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"

```

BASE COUNT	152 a	113 c	119 g	168 t	"/note="Vector: pbk-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN					
Query Match		2.2%	Score 41.4;	DB 12;	Length 552;
Best Local Similarity		49.3%	Pred. No. 9.1;		
Matches 108;	Conservative	0;	Mismatches 111;	Indels	0;
Gaps					
Qy	383	AATCTTTTGCTTATCTGGGCA	CTTACAGTATTTGGATGCC	CACTACCTTTGCAGCA	442
Db	62	AATATATTGGTTATGGGTGA	AGGTGAAGCAGGGGTGTC	CAAAATGCTGATTTTGA	121
Qy	443	CTCGTGGGATTTGTTCTTG	CTATGGGGATGGCCGTA	NGCAAAAGTCTTGATTTG	GA 502
Db	122	CTTGCAAGATTTTATCAAG	CTCCTTGGAAGCATTAT	CTGAANAAGTGTATTTGA	CT 181
Qy	503	AGAAATCCAGAGGAATTT	TATTTATGTCTAAAGT	CTTAAAAATCTGTAGAAA	AAGATAT 562
Db	182	ATTTGATACCGTGACCT	CGAATGCTTTTGGTGA	AAAGCACTATACAGTCA	ATTGAT 241
Qy	553	ACCAAGGCTTTTGGAGC	ATTTTTGATTTCTAAT	CTTGACT 601	
Db	242	ATGTGGCGCTTGGATGT	ATTATTTTTCGAGCT	TTTGTACT 280	

RESULT 38					
B0349244/c					
LOCUS	B0349244	676 bp	mRNA	linear	EST 06-MAR-2002
DEFINITION	B0349244 Dictyostelium discoideum cDNA library, AF Dictyostelium discoideum cDNA clone dda35123 3', mRNA sequence.				
ACCESSION	U710044				

SOURCE ORGANISM	REFERENCE AUTHORS	TITLE
Dicystostelium discoidium.	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium. 1 (bases 1 to 676)	Unishihara, H., Tanaka, Y., Kohara, Y., and Shin-i, T. Full length cDNA of Dicystostelium discoidium at the aggregation stage
JOURNAL COMMENT	Unpublished (2002) Contact : Tadaasu Shin-i	

Center for Genetic Resource Information  
National Institute of Genetics  
111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: [tshini@genes.nig.ac.jp](mailto:tshini@genes.nig.ac.jp).

FEATURES	SOURCE	LOCATION/Qualifiers
	1. 676	
		/organism="Dictyostelium discoideum"
		/strain="AX4"
		/db_xref="taxon:44689"
		/clone="dda35123"
		/clone_lib="Dictyostelium discoideum CDNA library, AF"
		/sex="mat A"
		/dev_stage="Aggregation stage"
BASE COUNT	198 a	125 c 120 g 232 t 1 others
ORIGIN		

	Query Match	2.2%	Score 41.4;	DB 13;	Length 676;
	Best Local Similarity	50.0%	Pred. No. 9.8;		
	Matches 102;	Conservative 0;	Mismatches 102;	Indels 0;	Gaps 0;
Qy	836	GGAAGTCTTTTCTTTTAGTGGTGGCTCTCGGGTTGAGCCTGGAATTCGCTTTTG	895		
Db	210	GGTAAAGTTACTTGTCATTGGTATCGTTCCAAAGATTCGAAATCTCAAAAATTAACT	151		
Qy	896	GGAATGATTTTAAAGAGGAGTATGCTTTTACCTTTATTCAAAAGACATGGCATCAGC	955		
Db	150	GCAATTCGTTGGCTTTGTAAAGTTATCCGAATTTTAGTGGTAAAGGTGGTGTAAATT	91		
Qy	956	GATGTGCTCAATAGCGTGCGAAGTTGTGCATTAACACTACAGAAAGCTGCTCTTTCTCT	1015		









**ACCESSION** BF295860  
**VERSION** BF295860.1 GI:13946305  
**KEYWORDS** EST.  
**SOURCE** Plasmodium berghei.  
**ORGANISM** Plasmodium berghei.  
**REFERENCE** 1 (bases 1 to 629)  
**AUTHORS** Carlton, J.M.-R. and Dame, J.B.  
**TITLE** The Plasmodium vivax and P. berghei gene sequence tag projects  
**JOURNAL** Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)  
**COMMENT** Contact: Dame JB  
 Department of Pathobiology, College of Veterinary Medicine  
 University of Florida  
 2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611-0880, USA  
 Tel: 352 392 4700  
 Fax: 352 392 9704  
 Email: damej@mail.vetmed.ufl.edu  
 Seq primer: T3.

**FEATURES** Location/Qualifiers  
 source 1..629  
 /organism="Plasmodium berghei"  
 /strain="ANKA clone Hp (gametocyte producer)"  
 /db\_xref="taxon:5821"  
 /clone\_lib="Pb cDNA #17, Tommaso Pace, Marta Ponzi, and Clara Frontali"  
 /dev\_stage="asynchronous blood stage"  
 /lab\_host="Wistar rats"  
 /note="Vector: pBluescript II vector DNA, excised from Lambda ZAP II.; Site 1: EcoRI; Site 2: XhoI; Total RNA was extracted from asynchronous blood stage forms of the cloned ANKA isolate of P. berghei, grown in wistar rats to 30% parasitemia and 2-5% gametocytemia. Contaminating host cell cells had previously been removed and final host cell contamination estimated to be approximately 5%. PolyA+ RNA was extracted and reverse transcribed using an oligo dT-XhoI primer (Lambda ZAP II cDNA cloning kit, Stratagene). Second strand cDNA was made following the manufacturer's protocol. EcoRI adaptors were ligated to the cDNA, and fragments were ligated into EcoRI/XhoI digested vector."

**BASE COUNT** 251 a 47 c 76 g 252 t 3 others  
**ORIGIN**

Query Match 2.2%; Score 41.2; DB 12; Length 629;  
 Best Local Similarity 49.3%; Pred. No. 11;  
 Matches 103; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

Qy 1606 AGTTTGTAAATGCTTTGTTATAGCGGCTCCTGCTGCTTAATTTGCAATTTATAT 1665  
 ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 616 AGTTTTTTTTTTTTTTTAAATTTTGGTTTGGTAAATTTGNAATGGAAAT 557  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 1666 GACCATAGGATCTTCTAGGAACCTTATCGTCTCTTATATTGCACCACTCTGTTGT 1725  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 556 AATATACAAATCATATAGCTTATTAATTTCTTATATTTCTTACCCCTTAATTTT 497  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 1726 GTTATGTCGTAAGAAATTCGCTCAAAATAGTACCGTTAAATCTTAATCAACGTGT 1785  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 496 TTATACCATCACTACATATTTATTTTAAATAAAATGCAATAACAATAAAATTTG 437  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 1786 AGCATATAAAATCTCTTTGGGACTTT 1814  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 436 CATAAAAAAACTAGTCCCTCTTTTCTTT 408  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

**RESULT** 44  
**CNS0111D/c**  
**LOCUS** CNS0111D 750 bp DNA linear GSS 26-JUL-1999  
**DEFINITION** Drosophila melanogaster genome survey sequence T7 end of BAC BACN06D21 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
**ACCESSION** AL100303  
**VERSION** AL100303.1 GI:5611914  
**KEYWORDS** GSS.

**SOURCE** Drosophila melanogaster.  
**ORGANISM** Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
**REFERENCE** 1 (bases 1 to 750)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
**COMMENT** Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

**FEATURES** Location/Qualifiers  
 source 1..750  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone="BACN06D21"  
 /clone\_lib="DrosBAC"  
 /plasmid="pBelOBAC11"  
 /notes="end : T7"

**BASE COUNT** 309 a 104 c 111 g 102 t 124 others  
**ORIGIN**

Query Match 2.2%; Score 41.2; DB 17; Length 750;  
 Best Local Similarity 34.9%; Pred. No. 11;  
 Matches 97; Conservative 44; Mismatches 137; Indels 0; Gaps 0;

Qy 621 TTCTTTCTCTAGATACAGGCGCTATTAAAGGTTGCTTTCACATTTGATTTAGGAA 680  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 386 TTYTTTWTWKTTTWTCTAAGTTKTYTAGTTTCTTATTTCTTAAKTKAKTTT 327  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 681 TTTTCTCTCAATGTTTACGGCTCTTTTCATGACTAAATTTTCTTCATGCTGGATGA 740  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 326 TTTTCTCTKAGYTTTWTGTTTCTTTTCTGTAATKTKTKTKTTTCTTTTCKM 267  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 741 ATAAGACCAACATACACAGTTGCATATGATGAATAGTTCTGCGGGATAAAGCATGAT 800  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 266 KTAATTTTKYTTAATKATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 207  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 801 TCTTGAGAGGATCAAAAACCTTTGGGCTGTTTCTGGAAGTGTCTTTCTTTAGTTCG 860  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 206 WTTTWTAKTTTTCGAKKTTCTTTTWTYTKGTWKYKTTTGTGTTTWTWKTKT 147  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Qy 861 TTGCTCTCGGTTTGGAGCCTGGAATCCGTTTGGGA 898  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 146 CTTTWTWKTTATTGCGCTCAAGTCCAGTTGTMGCA 109  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

**RESULT** 45  
**C92475/c**  
**LOCUS** C92475 436 bp mRNA linear EST 12-JUL-1999  
**DEFINITION** C92475 Dictyostellium discoideum SS (H.urushihara) Dictyostellium discoideum CDNA clone SSE617, mRNA sequence.  
**ACCESSION** C92475  
**VERSION** C92475.1 GI:3074351  
**KEYWORDS** EST.  
**SOURCE** Dictyostellium discoideum.  
**ORGANISM** Dictyostellium discoideum  
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.  
**REFERENCE** 1 (bases 1 to 436)  
**AUTHORS** Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M., Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H., Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.  
**TITLE** The dictyostellium developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development



















C 959	16	0.9	3983	22	AAH54945	S. epidermidis gen
960	16	0.9	4000	23	AA864530	DNA encoding novel
961	16	0.9	4006	21	AA16145	Human prostate can
C 962	16	0.9	4024	24	AA594815	Human DNA sequence
C 963	16	0.9	4077	22	AA169450	Human tumor-asso
C 964	16	0.9	4097	22	AAH26923	Human G protein as
965	16	0.9	4142	24	AA276627	Human colon specifi
966	16	0.9	4149	23	ABV24665	Human prostate exp
967	16	0.9	4229	23	AB119610	Drosophila melanog
968	16	0.9	4285	21	AA61276	Human secreted pro
C 969	16	0.9	4307	23	AB106079	Drosophila melanog
C 970	16	0.9	4328	13	AA28758	Partial sequence o
C 971	16	0.9	4338	22	AA126425	Human breast cance
972	16	0.9	4338	22	AA126425	Human breast cance
C 973	16	0.9	4341	23	AB120386	Drosophila melanog
974	16	0.9	4370	23	AB108450	Drosophila melanog
C 975	16	0.9	4389	23	AB104744	Drosophila melanog
C 976	16	0.9	4390	23	AB105629	Drosophila melanog
C 977	16	0.9	4428	23	AB130129	Drosophila melanog
978	16	0.9	4465	22	AA61044	Human cancer agent
C 979	16	0.9	4547	23	AB108012	Drosophila melanog
C 980	16	0.9	4576	23	AA89475	DNA encoding novel
C 981	16	0.9	4635	23	AB130220	Drosophila melanog
C 982	16	0.9	4646	22	AAH72849	Human cervical can
983	16	0.9	4651	15	AA072913	Human receptor-tyr
C 984	16	0.9	4774	23	AB113187	Drosophila melanog
C 985	16	0.9	4776	23	AB117072	Drosophila melanog
C 986	16	0.9	4814	23	AB118054	Drosophila melanog
C 987	16	0.9	4823	23	AA592612	DNA encoding novel
C 988	16	0.9	4825	23	AB128874	Drosophila melanog
C 989	16	0.9	4887	23	AB110090	Drosophila melanog
C 990	16	0.9	5019	21	AA65709	Arabidopsis thalia
C 991	16	0.9	5034	22	AB16829	Human nervous syst
C 992	16	0.9	5034	22	AAK82034	Human immune/haema
C 993	16	0.9	5036	22	AB16827	Human nervous syst
C 994	16	0.9	5036	22	AB16828	Human nervous syst
C 995	16	0.9	5036	22	AAK82030	Human immune/haema
C 996	16	0.9	5036	22	AAK82033	Human immune/haema
C 997	16	0.9	5046	22	AAH48716	T. thermophila tri
C 998	16	0.9	5080	23	ABK42310	Genomic sequence #
C 999	16	0.9	5080	23	ABK42311	Genomic sequence #
C1000	16	0.9	5111	19	AAV44248	Lettuce resistance

## ALIGNMENTS

RESULT 1  
AA50030 standard; DNA; 1864 BP.

AC AA50030;  
XX 10-OCT-2000 (first entry)  
DT DNA encoding Chlamydia pneumoniae antigen CPN100686 RY 54.  
DE CPN100686 RY 54; antigen; infection; diagnosis; therapy; vaccine;  
KM outer membrane protein, de.  
XX Chlamydia pneumoniae.  
OS  
XX  
FH Key Location/Qualifiers  
FT CDS 101..1759 /\*tag= a  
PN WO200039158-A1.  
XX  
PD 06-JUL-2000.  
XX  
PF 23-DEC-1999; 99WO-CA01230.  
XX  
PR 23-DEC-1998; 98US-0113280.

PR	23-DEC-1998;	98US-0113281.
PR	23-DEC-1998;	98US-0113282.
PR	23-DEC-1998;	98US-0113283.
PR	23-DEC-1998;	98US-0113284.
PR	23-DEC-1998;	98US-0113285.
PR	23-DEC-1998;	98US-0113385.
PR	28-DEC-1998;	98US-0114050.
PR	28-DEC-1998;	98US-0114056.
PR	28-DEC-1998;	98US-0114057.
PR	28-DEC-1998;	98US-0114058.
PR	28-DEC-1998;	98US-0114059.
PR	28-DEC-1998;	98US-0114061.
PA	(CONN-) CONNAUGHT LAB LTD.	
XX	Muridin AD, Oomen RP, Wang J;	
PI	WPI; 2000-452369/39.	
DR	P-PSDB; AA95543.	
XX	Novel Chlamydia polynucleotides and polypeptides useful for diagnosis,	
PT	prevention and treatment of Chlamydia infection in mammals -	
XX	Claim 2(a); Page 55-58; 215pp; English.	
PS	The present sequence is that of Chlamydia pneumoniae genomic DNA	
CC	including an open reading frame that codes for CPN100686 RY 54 (see	
CC	AA95543), a putative 93 kDa outer membrane protein. It is an example	
CC	of C. pneumoniae polynucleotide molecules of the invention (see	
CC	AA50030-42) that encode antigenic polypeptides (see AA95543-55) useful	
CC	in the diagnosis, treatment and prevention of Chlamydia infection.	
CC	The polynucleotides can be utilised: in the recombinant production	
CC	of Chlamydia antigens using transformed unicellular host cells; in	
CC	vaccines or live vaccine vectors; in naked form or formulated with	
CC	a delivery vehicle for therapy and prophylaxis of Chlamydia	
CC	infection; and as probes or primers for diagnosis of Chlamydia	
CC	infection.	
XX	Sequence 1864 BP; 498 A; 346 C; 401 G; 619 T; 0 other;	
SQ		
Query Match	100.0%; Score 1864; DB 21; Length 1864;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 1864; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 ATGACACTCCGCATTTGTTCAGAGGGGATCAGCGGACCTGCTAAATGACATATTTCTGC	60
DB	1 ATGACACTCCGCATTTGTTCAGAGGGGATCAGCGGACCTGCTAAATTTCTGC	60
QY	61 AAACCGTGAATGGCGTATGCTGATGATTAAGGTTATATGTCAGAGCCCTATTTT	120
DB	61 AAACCGTGAATGGCGTATGCTGATGATTAAGGTTATATGTCAGAGCCCTATTTT	120
QY	121 AAACGTCCTGTAATAATCATGCCAGTGTCTCAGGGAATTTTACCACCGCTAAGTAG	180
DB	121 AAACGTCCTGTAATAATCATGCCAGTGTCTCAGGGAATTTTACCACCGCTAAGTAG	180
QY	181 CAACTGCTCTGATTTAAATCTGAGAGGATGCTTTTGTCCGAGGTTCTCAGTGA	240
DB	181 CAACTGCTCTGATTTAAATCTGAGAGGATGCTTTTGTCCGAGGTTCTCAGTGA	240
QY	241 AGAGAGCATCTCTGATTTGAGGAAAAAACAATGATGACAGGAGATTATCTCAGATG	300
DB	241 AGAGAGCATCTCTGATTTGAGGAAAAAACAATGATGACAGGAGATTATCTCAGATG	300
QY	301 CTGTGGCTTGGCAATGCTTATTTGATGAGCGTATATATATAGATTGAGAGCGTCAT	360
DB	301 CTGTGGCTTGGCAATGCTTATTTGATGAGCGTATATATATAGATTGAGAGCGTCAT	360
QY	361 CGCTTGGAGAGCTGTTCTTCTGAATCTTTTGTATCTGGGACGCTCTACAGTATTTGA	420
DB	361 CGCTTGGAGAGCTGTTCTTCTGAATCTTTTGTATCTGGGACGCTCTACAGTATTTGA	420
QY	421 TGGCCACTACCTTGTGAGAGACTCGTGGAGTTGTTCTTGCTATGGGATGGCCGTAGA	480



PR 21-JUL-1997; 97US-0897798.  
 PR 01-JUL-1997; 97US-0051407.  
 XX  
 XX (INGH/) INGHAM P W.  
 PA (ONTO-) ONTOGENY INC.  
 PA (VHEU/) VAN DEN HEUVEL M.  
 XX  
 XX WPI; 1999-105996/09.  
 DR P-PSDB; AAW81066.  
 XX  
 PT New isolated smoothened genes - used to develop products for  
 PT treating e.g. cancers, neurodegenerative disorders, nervous system  
 PT injury, immunological diseases or infections  
 XX  
 PS Claim 12; Pages 93-94; 107pp; English.

CC This is the nucleotide sequence of the novel chicken smoothened  
 CC gene used in the method of the invention to produce smoothened  
 CC polypeptides (SPs). The SPs can modulate at least one of  
 CC proliferation, differentiation or survival of a cell which  
 CC expresses the SP. The products can be used to modulate  
 CC spermatogenesis, osteogenesis, chondrogenesis or neuronal cell  
 CC differentiation or to enhance survival of neuronal cells, e.g. to  
 CC prevent apoptosis. They can be used for treating e.g. cancer,  
 CC degenerative disorders marked by loss of particular cell-types,  
 CC apoptosis, neoplastic and/or hyperplastic disorders. In particular  
 CC they can be used to treat neurological conditions derived from  
 CC acute, subacute, or chronic injury to the nervous system, aging of  
 CC the nervous system chronic neurodegenerative diseases of the  
 CC nervous system, chronic immunological diseases of the nervous  
 CC system or affecting the nervous system. They can also treat be used  
 CC to CNS trauma infection, infection (such as viral infection with  
 CC varicella-zoster), metabolic disease, nutritional deficiency,  
 CC toxic agents (such as cisplatin treatment), chronic pain syndromes  
 CC or in nerve prostheses for the repair of central and peripheral  
 CC nerve damage, malignant gliomas, medulloblastomas, neuroectodermal  
 CC tumours, ependymomas and for liver repair and regeneration of lung  
 CC tissue in the treatment of emphysema, or for enhancing bond  
 CC formation, tissue formation or fertility. The products can also be  
 CC used for detection, diagnosis and drug screening.

XX Sequence 3256 BP; 594 A; 1015 C; 953 G; 669 T; 25 other;

Query Match 1.2%; Score 22; DB 20; Length 3256;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1277 GCTTTGGCAATCATCTTGTCT 1298  
 DB 1697 GCTTTGGCAATCATCTTGTCT 1676

RESULT 3  
 ABQ33748/c  
 ID ABQ33748 standard; DNA; 570 BP.  
 AC  
 XX  
 AC ABQ33748;  
 XX

DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 20339.  
 XX  
 XX Human: cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KM drug; side effect; cancer; central nervous system; cardiovascular;  
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KM SNP; cell differentiation; ds.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 PN WO200218632-A2.  
 XX  
 PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.  
 XX  
 XX 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 XX (EPIC-) EPIGENOMICS AG.  
 PA  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX  
 XX WPI; 2002-371829/40.  
 DR  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert a  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ33410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX Sequence 570 BP; 93 A; 53 C; 218 G; 206 T; 0 other;

Query Match 1.1%; Score 21; DB 24; Length 570;  
 Best Local Similarity 100.0%; Pred. No. 3.7;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 GAAACTCTAAGCAAGCA 1205  
 DB 287 GAAACTCTAAGCAAGCA 267

RESULT 4  
 ABQ33749  
 ID ABQ33749 standard; DNA; 570 BP.  
 AC  
 XX  
 AC ABQ33749;  
 XX

DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 20340.  
 XX  
 XX Human: cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KM drug; side effect; cancer; central nervous system; cardiovascular;  
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KM SNP; cell differentiation; ds.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 PN WO200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP10074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.  
 XX (EPiG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 DR - WPI; 2002-371829/40.  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 XX  
 SQ Sequence 570 BP; 206 A; 218 C; 53 G; 93 T; 0 other;  
 Query Match 1.1%; Score 21; DB 24; Length 570;  
 Best Local Similarity 100.0%; Pred. No. 3.7;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1185 GAAACTCTAAACGAAACGCAA 1205  
 DB 284 GAAACTCTAAACGAAACGCAA 304  
 RESULT 5  
 ABQ32730/C  
 ID ABQ32730 standard; DNA; 573 BP.  
 XX  
 AC ABQ32730;  
 XX  
 DT 12-JUL-2002 (first entry)  
 DE  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 19321.  
 XX  
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EPI0074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPiG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX

XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX  
 DR WPI; 2002-371829/40.  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 XX  
 SQ Sequence 573 BP; 82 A; 49 C; 229 G; 207 T; 6 other;  
 Query Match 1.1%; Score 21; DB 24; Length 573;  
 Best Local Similarity 100.0%; Pred. No. 3.7;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1185 GAAACTCTAAACGAAACGCAA 1205  
 DB 200 GAAACTCTAAACGAAACGCAA 180  
 RESULT 6  
 ABQ32731  
 ID ABQ32731 standard; DNA; 573 BP.  
 XX  
 AC ABQ32731;  
 XX  
 DT 12-JUL-2002 (first entry)  
 DE  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 19322.  
 XX  
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EPI0074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPiG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX

DR WPI; 2002-371829/40.

XX  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA -

PS Claim 12; 56pp + Sequence listing; 56pp; German.

XX  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ1410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

CC  
XX  
SQ Sequence 573 BP; 207 A; 229 C; 49 G; 82 T; 6 other;

Query Match 1.1%; Score 21; DB 24; Length 573;  
Best Local Similarity 100.0%; Pred.No.3.7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 GAAACTCTAAACGAAACGCAA 1205  
DB 374 GAAACTCTAAACGAAACGCAA 394

RESULT 7  
AA97008/C  
ID AAX97008 standard; DNA; 20 BP.

XX  
XX AAX97008;

DT 13-SEP-1999 (first entry)

XX  
XX PCR primer used to amplify an ORF of Chlamydia pneumoniae.

DE  
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope; PCR primer; ss.

XX  
OS Synthetic.  
OS Chlamydia pneumoniae.

XX  
PN WO927105-A2.

PD 03-JUN-1999.

XX  
PF 20-NOV-1998; 98WO-IB01890.

XX  
PR 04-NOV-1998; 98US-0107078.  
PR 21-NOV-1997; 97FR-0014673.

XX  
PA (GEST ) GENSET.

XX  
PI Griffais R;

XX  
XX WPI; 1999-357842/30.

PT Genome sequence of Chlamydia pneumoniae

XX  
XX Page 1870; Disclosure; 1912pp; English.

PS  
XX AAX91991-X97517 represent PCR primers used to amplify open reading  
CC frames and other nucleic acid sequences from the genome of  
CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory  
CC disease such as pneumonia and bronchitis and is thought to be a  
CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent  
CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded  
CC by the open reading frames of the C. pneumoniae genome (see AAY34584-  
CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors  
CC containing C. pneumoniae nucleotide sequences can also be used as  
CC immunogenic compositions, especially where the vector directs the  
CC expression of a neutralising epitope of C. pneumoniae.

XX  
SQ Sequence 20 BP; 5 A; 6 C; 4 G; 5 T; 0 other;

Query Match 1.1%; Score 20; DB 20; Length 20;  
Best Local Similarity 100.0%; Pred.No.11;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 ATTCTGCAAAACCGTGATGG 73  
DB 20 ATTCTGCAAAACCGTGATGG 1

RESULT 8  
AAX95807  
ID AAX95807 standard; DNA; 20 BP.

XX  
XX AAX95807;

DT 13-SEP-1999 (first entry)

XX  
XX PCR primer used to amplify an ORF of Chlamydia pneumoniae.

DE  
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope; PCR primer; ss.

XX  
OS Synthetic.  
OS Chlamydia pneumoniae.

XX  
PN WO927105-A2.

PD 03-JUN-1999.

XX  
PF 20-NOV-1998; 98WO-IB01890.

XX  
PR 04-NOV-1998; 98US-0107078.  
PR 21-NOV-1997; 97FR-0014673.

XX  
PA (GEST ) GENSET.

XX  
PI Griffais R;

XX  
XX WPI; 1999-357842/30.

PT Genome sequence of Chlamydia pneumoniae

XX  
PS Page 1777; Disclosure; 1912pp; English.

XX  
XX AAX91991-X97517 represent PCR primers used to amplify open reading  
CC frames and other nucleic acid sequences from the genome of  
CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory  
CC disease such as pneumonia and bronchitis and is thought to be a  
CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent  
CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded  
CC by the open reading frames of the C. pneumoniae genome (see AAY34584-  
CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors  
CC containing C. pneumoniae nucleotide sequences can also be used as  
CC immunogenic compositions, especially where the vector directs the  
CC expression of a neutralising epitope of C. pneumoniae.

ID	AAK93334 standard; DNA; 20 BP.
XX	
AC	AAK93334;
XX	
DT	13-SEP-1999 (first entry)
XX	
DE	PCR primer used to amplify an ORF of Chlamydia pneumoniae.
XX	
KW	Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KM	sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW	vaccine; neutralising epitope; PCR primer; ss.
XX	
OS	Synthetic.
OS	Chlamydia pneumoniae.
XX	
PN	WQ927105-A2.
XX	
PD	03-JUN-1999.
XX	
PX	20-NOV-1998; 98WO-IB01890.
XX	
PR	04-NOV-1998; 98US-0107078.
PR	21-NOV-1997; 97FK-0014673.
XX	
PA	(GEST ) GENSET.
XX	
PI	Griffais R;
XX	
DR	WPI; 1999-357842/30.
XX	
PT	Genome sequence of Chlamydia pneumoniae
PS	Page 1581; Disclosure; 1912pp; English.
XX	
CC	AAK91991-X97517 represent PCR primers used to amplify open reading
CC	frames and other nucleic acid sequences from the genome of
CC	Chlamydia pneumoniae (see AAK91990). C. pneumoniae causes respiratory
CC	disease such as pneumonia and bronchitis and is thought to be a
CC	contributing factor in heart disease, sarcoidosis, sinusitis, purulent
CC	otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
CC	by the open reading frames of the C. pneumoniae genome (see AAK91990-
CC	AAK91991) can be used in immunogenic compositions as vaccines. Vectors
CC	containing C. pneumoniae nucleotide sequences can also be used as
CC	immunogenic compositions, especially where the vector directs the
CC	expression of a neutralising epitope of C. pneumoniae.
XX	
SC	Sequence 20 BP; 5 A; 6 C; 4 G; 5 T; 0 other;
	Query Match 1.1%; Score 20; DB 20; Length 20;
	Best Local Similarity 100.0%; Pred No. 11;
	Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	953 AGCGATGGTCTCAAAATGCC 972
Db	20 AGCGATGGTCTCAAAATGCC 1
RESULT 11	
AAC05080	
ID	AAC05080 standard; cDNA; 236 BP.
XX	
AC	AAC05080;
XX	
DT	06-OCT-2000 (first entry)
XX	
DE	Human secreted protein 5' EST, SEQ ID NO: 9155.
XX	
KW	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW	gene therapy; chromosome mapping; ss.
XX	
OS	Homo sapiens.
XX	
PN	RP1013401-A2

```

XX 06-SEP-2000.
PD New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PF diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX 26-FEB-1999; 99US-0122487.
XX (GEST ) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 9155; 71bp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs with genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX
XX Sequence 236 BP; 69 A; 51 C; 43 G; 73 T; 0 other;
XX
XX Query Match 1.0%; Score 19; DB 21; Length 236;
XX Best Local Similarity 100.0%; Pred. No. 39;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 681 TTTCTCTTCATGTTTAC 699
DB 113 TTTCTCTTCATGTTTAC 131

```

RESULT 12  
AAC03937 standard; cDNA; 420 BP.

XX AAC03937;  
XX  
XX 06-OCT-2000 (first entry)  
XX  
XX Human secreted protein 5' EST, SEQ ID NO: 3935.  
XX  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping; ss.  
XX  
XX Homo sapiens.  
XX  
XX EP1033401-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 21-FEB-2000; 2000EP-0200610.  
XX  
XX 26-FEB-1999; 99US-0122487.  
XX  
XX (GEST ) GENSET.  
XX  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX WPI; 2000-500381/45.  
XX

```

DR P-PSDB; AAC03931.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 3935; 71bp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. An ORF has been identified within the
XX sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
XX derived from 30 different tissues. EST sequences usually correspond
XX mainly to the 3' untranslated region (UTR) of the mRNA because they are
XX often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
XX well suited for isolating cDNA sequences derived from the 5' ends of
XX mRNAs and even in those cases where longer cDNA sequences have been
XX obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
XX mRNAs with intact 5' ends and can therefore be used to obtain full length
XX cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
XX gene therapy and chromosome mapping procedures. They are used to obtain
XX upstream regulatory sequences and to design expression and secretion
XX vectors.
XX
XX Sequence 420 BP; 100 A; 112 C; 107 G; 101 T; 0 other;
XX
XX Query Match 1.0%; Score 19; DB 21; Length 420;
XX Best Local Similarity 100.0%; Pred. No. 40;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 370 AGCTGTTCTTCGATCTT 388
DB 376 AGCTGTTCTTCGATCTT 394

```

RESULT 13  
AAK03932/c  
ID AAK03932 standard; DNA; 439 BP.

XX AAK03932;  
XX  
XX 05-NOV-2001 (first entry)  
XX  
XX Human brain expressed single exon probe SEQ ID NO: 3923.  
XX  
XX Human; brain expressed exon; gene expression analysis; probe;  
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200157275-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00667.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX  
XX 26-MAY-2000; 2000US-0207456.  
XX  
XX 30-JUN-2000; 2000US-0608408.  
XX  
XX 03-AUG-2000; 2000US-0632366.  
XX  
XX 21-SEP-2000; 2000US-0234687.  
XX  
XX 27-SEP-2000; 2000US-0236359.  
XX  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -  
XX



PS Example 4; SEQ ID NO: 3923; 650pp + Sequence Listing; English.

The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention.

SQ Sequence 439 BP; 126 A; 97 C; 89 G; 127 T; 0 other;

Query Match 1.0%; Score 19; DB 22; Length 439;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 19; Conservative 0; Mismatches 0; Indels

Qy 370 AGCTGTTCTTCTGAATCTT 388  
|||  
Db 412 AGCTGTTCTTCTGAATCTT 394

RESULT 14  
AAI13982/c  
ID AAI13982 standard; DNA; 439 BP.

AC AAI13982;

DT 12-OCT-2001 (first entry)

Probe #3915 for gene expression analysis in human cervical cell sample.

Probe; human; microarray; gene expression; cervical epithelial cell; cervical cancer; ss.

OS Homo sapiens.

PN WO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR	04-FEB-2000;	2000US-0180312.
PR	26-MAY-2000;	2000US-0207456.
PR	30-JUN-2000;	2000US-0608408.
PR	03-AUG-2000;	2000US-0632366.
PR	21-SEP-2000;	2000US-0234687.
PR	27-SEP-2000;	2000US-0236359.
PR	04-OCT-2000;	2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI	Penn SG,	Hanzel DK,	Chen W,	Rank DR:

DR WPI; 2001-488901/53.

PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human cervical epithelial cells -  
PT

PS Claim 25: SEO ID No 3915: 487pp: English.

The present invention relates to human single exon nucleic acid probes (SENPs). The present sequence is one such probe. The SENPs are derived from human HeLa cells. The SENPs can be used to produce a single exon microarray, which can be used for measuring human gene expression in a sample derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at <ftp.wipo.int/pub/published> pct sequences.

SQ Sequence 439 BP; 126 A; 97 C; 89 G; 127 T; 0 other;

Query Match 1.0%; Score 19; DB 22; Length 439;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 19; Conservative 0; Mismatches 0; Indels

Qy 370 AGCTGTTCTTCTGAATCTT 388  
|||  
Db 412 AGCTGTTCTTCTGAATCTT 394

RESULT 15  
AAI03846/c  
ID AAI03846 standard; DNA; 439 BP.

AAI03846;

DT 09-OCT-2001 (first entry)

Probe #3837 used to measure gene expression in human breast sample.

probe; human; breast disease; breast cancer; development disorder; ss;  
KW  
KW  
KW  
inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

WO200157270-A2

09-AUG-2001.

AA 29-JAN-2001; 2001WO-US00661.

PR	04-FEB-2000;	2000US-0180312.
PR	26-MAY-2000;	2000US-0207456
PR	30-JUN-2000;	2000US-0608408
PR	03-AUG-2000;	2000US-0632366.
PR	21-SEP-2000;	2000US-0234687.
PR	27-SEP-2000;	2000US-0236359.
PR	04-OCT-2000;	2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG. Hanzel DK. Chen W. Rank DR:

WPI: 2001-476286/51.

Novel single exon nucleic acid probe used to measuring gene expression in a human breast -

PS Claim 25: SEO ID No 3837: 322pp: English: XX

The present invention relates to novel single exon nucleic acid probes. The present sequence is one such probe. The probes are useful for measuring human gene expression in a human breast sample, where the probe hybridises at high stringency to a nucleic acid expressed in the human breast. The probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human breast, particularly those diseases with polygenic aetiology. The diseases include: breast cancer, disorders of development, inflammatory diseases of the breast, fibrocystic changes, proliferative breast disease and non-carcinoma tumours.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published/pct/sequences](http://ftp.wipo.int/pub/published/pct/sequences).

Sequence 439 BP: 126 A: 97 C: 89 G: 127 T: 0 other:

Query Match 1.0%; Score 19; DB 22; Length 439;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 19: Conservative 0; Mismatches 0: Indels

Qy 370 AGCTGTTCTTCTGAATCTT 388  
|||  
Db 412 AGCTGTTCTTCTGAATCTT 394

```

RESULT 16
AAK16682/c
ID AAK16682 standard; DNA; 445 BP.
XX
XX
AC AAK16682;
XX
XX
DT 05-NOV-2001 (first entry)
XX
XX
DE Human brain expressed single exon probe SEQ ID NO: 16673.
XX
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200157275-A2.
XX
XX
PD 09-AUG-2001.
XX
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0603408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX
PT WPI; 2001-483446/52.
XX
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
XX
PS Example 4; SEQ ID NO: 15673; 650bp + Sequence Listing; English.
XX
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
XX
SQ Sequence 445 BP; 119 A; 99 C; 93 G; 134 T; 0 other;
XX
XX
Query Match 1.0%; Score 19; DB 22; Length 445;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX
QY 370 AGCTGTTCTTCTGAATCTT 388
DB 286 AGCTGTTCTTCTGAATCTT 268
XX
XX
RESULT 17
AAI23198/c
ID AAI23198 standard; DNA; 445 BP.
XX
XX
AC AAI23198;
XX
XX
DT 12-OCT-2001 (first entry)
XX
XX
DE Probe #1131 for gene expression analysis in human cervical cell sample.
XX
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX

```

```

KW cervical cancer; ss.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200157278-A2.
XX
XX
PD 09-AUG-2001.
XX
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX
PT WPI; 2001-488901/53.
XX
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX
PS Claim 25; SEQ ID No 13131; 487bp; English.
XX
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 445 BP; 119 A; 99 C; 93 G; 134 T; 0 other;
XX
XX
Query Match 1.0%; Score 19; DB 22; Length 445;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX
QY 370 AGCTGTTCTTCTGAATCTT 388
DB 286 AGCTGTTCTTCTGAATCTT 268
XX
XX
RESULT 18
AAI08848/c
ID AAI08848 standard; DNA; 445 BP.
XX
XX
AC AAI08848;
XX
XX
DT 09-OCT-2001 (first entry)
XX
XX
DE Probe #8839 used to measure gene expression in human breast sample.
XX
XX
KW Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200157270-A2.
XX
XX
PD 09-AUG-2001.
XX
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
XX
PR 04-FEB-2000; 2000US-0180312.
XX

```

PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-476286/51.  
XX Novel single exon nucleic acid probe used to measuring gene expression  
PT in a human breast -  
XX Claim 25; SEQ ID No 8839; 322pp; English.  
XX The present invention relates to novel single exon nucleic acid probes.  
CC The present sequence is one such probe. The probes are useful for  
CC measuring human gene expression in a human breast sample, where the probe  
CC hybridises at high stringency to a nucleic acid expressed in the human  
CC breast. The probes are useful for predicting, diagnosing, grading,  
CC staging, monitoring and prognosing diseases of the human breast,  
CC particularly those diseases with polygenic aetiology. The diseases  
CC include: breast cancer, disorders of development, inflammatory diseases  
CC of the breast, fibrocystic changes, proliferative breast disease and  
CC non-carcinoma tumours.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 445 BP; 119 A; 99 C; 93 G; 134 T; 0 other;  
Query Match 1.0%; Score 19; DB 22; Length 445;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 370 AGCTGTTCTTCTGAATCTT 388  
Db 286 AGCTGTTCTTCTGAATCTT 268  
RESULT 19  
AAK50436/c  
ID AAK50436 standard; DNA; 469 BP.  
XX  
AC AAK50436;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 24993.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488900/53.  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
XX Example 4; SEQ ID NO: 24993; 658pp + Sequence Listing; English.  
XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention.  
XX  
SQ Sequence 469 BP; 125 A; 106 C; 98 G; 140 T; 0 other;  
Query Match 1.0%; Score 19; DB 22; Length 469;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 370 AGCTGTTCTTCTGAATCTT 388  
Db 310 AGCTGTTCTTCTGAATCTT 292  
RESULT 20  
AAK37567/c  
ID AAK37567 standard; DNA; 561 BP.  
XX  
AC AAK37567;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 12124.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PN Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488900/53.  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
XX Example 4; SEQ ID NO: 12124; 658pp + Sequence Listing; English.  
XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention.

CC the probes of the invention.  
 CC  
 XX  
 SQ Sequence 561 BP; 157 A; 123 C; 114 G; 167 T; 0 other;  
 Query Match 1.0%; Score 19; DB 22; Length 561;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 370 AGCTGTTCTTCGAACTT 388  
 |||||  
 DB 356 AGCTGTTCTTCGAACTT 338  
 RESULT 21  
 AAH08684  
 ID AAH08684 standard; cDNA; 812 BP.  
 XX  
 AC AAH08684;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA clone (5'-primer) SEQ ID NO:5519.  
 XX  
 KM Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 1; SEQ ID 5519; 2537bp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises at least 15 nucleotides, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 CC  
 XX  
 SQ Sequence 812 BP; 216 A; 188 C; 206 G; 198 T; 4 other;  
 Query Match 1.0%; Score 19; DB 22; Length 812;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 370 AGCTGTTCTTCGAACTT 388  
 |||||  
 DB 350 AGCTGTTCTTCGAACTT 368  
 RESULT 22  
 AAH05672  
 ID AAH05672 standard; cDNA; 827 BP.  
 XX  
 AC AAH05672;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA clone (5'-primer) SEQ ID NO:2507.  
 XX  
 KM Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 1; SEQ ID 2507; 2537bp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX  
 SQ Sequence 827 BP; 222 A; 192 C; 192 G; 208 T; 13 other;  
 Query Match 1.0%; Score 19; DB 22; Length 827;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCTGTTCTTCTGAATCTT 388  
 |||||  
 Db 358 AGCTGTTCTTCTGAATCTT 376

RESULT 23  
 ABK35236  
 ID ABK35236 standard; cDNA; 1150 BP.

XX  
 AC ABK35236;

XX  
 DT 08-MAY-2002 (first entry)

XX  
 DE Human cDNA encoding secreted protein #374.

XX Human; secreted protein; gene; ss; nutritional supplement; haemophilia;  
 KW viral infection; bacterial infection; fungal infection; diabetes; asthma;  
 KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;  
 KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;  
 KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;  
 KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;  
 KW tissue regeneration; wound healing; burn; haematopoiesis;  
 KW myeloid cell deficiency; lymphoid cell deficiency.

XX  
 OS Homo sapiens.

XX  
 PN WO200177288-A2.

XX  
 PD 18-OCT-2001.

XX  
 PF 29-MAR-2001; 2001WO-US10224.

XX  
 PR 06-APR-2000; 2000US-195582P.

XX  
 PA (GEMY ) GENETICS INST INC.

XX  
 PI Wong GG, Clark HF, Fachtel K, Agostino MJ, Howes SH, Resnick RJ;  
 PI Gulukota K, Graham JR;

XX  
 DR WPI; 2002-179321/23.

XX Five hundred and ninety two polynucleotides derived from a variety of  
 PT human tissue sources which encode secreted proteins, useful for  
 PT treating immune deficiencies and disorders such as autoimmune disorders

XX  
 PS Claim 1; Page 268; 372pp; English.

XX The invention relates to 592 polynucleotides which have been derived from  
 CC a variety of human tissue sources and which encode novel secreted  
 CC proteins. The polynucleotides can be used as probes for the  
 CC identification and isolation of full length cDNA and genomic DNA. The  
 CC polynucleotides and proteins can also be used as nutritional supplements.  
 CC The proteins are useful in the treatment of various immune deficiencies  
 CC and disorders such as viral infections, bacterial infections, fungal  
 CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple  
 CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions  
 CC and conditions (e.g. asthma). They are also useful for treating  
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's  
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),  
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also  
 CC useful for tissue regeneration, for wound healing and in the treatment of  
 CC burns, incisions and ulcers. The proteins are also useful for regulating

CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.  
 CC Sequences ABK34863-ABK35454 represent polynucleotides of the invention.

XX  
 SQ Sequence 1150 BP; 326 A; 247 C; 287 G; 290 T; 0 other;

Query Match 1.0%; Score 19; DB 24; Length 1150;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCTGTTCTTCTGAATCTT 388

|||||

Db 337 AGCTGTTCTTCTGAATCTT 355

RESULT 24

AAC81029

ID AAC81029 standard; cDNA; 1187 BP.

XX  
 AC AAC81029;

XX  
 DT 13-FEB-2001 (first entry)

XX  
 DE Human secreted protein cDNA sequence #2.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein; ss.

XX  
 OS Homo sapiens.

XX  
 PN WO200063230-A2.

XX  
 PD 26-OCT-2000.

XX  
 PF 23-MAR-2000; 2000WO-US07677.

XX  
 PR 26-MAR-1999; 99US-0126601.

XX  
 PR 17-SEP-1999; 99US-0154373.

XX  
 PR 14-JAN-2000; 2000US-0176064.

XX  
 PA (HUMA-) HUMAN GENOME SCI INC.

XX  
 PI Rosen CA, Ruben SM, Komatsoulis G;

XX  
 DR WPI; 2000-647515/62.

XX  
 DR P-PSDB; AAB45227.

XX New nucleic acid molecules encoding 49 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives -

XX  
 PS Claim 1; Page 340; 402pp; English.

XX The present invention relates to the isolation of genes encoding  
 CC 49 human secreted proteins. The genes can be used to  
 CC generate fusion proteins by linking to the gene for the human  
 CC immunoglobulin G Fc portion for increasing the stability of  
 CC the fusion protein as compared to the human protein only. The genes and  
 CC proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g. by protein or gene therapy. The genes are isolated  
 CC from a range of human tissues disclosed in the specification. The  
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;  
 CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound  
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;  
 CC and (f) infectious diseases such as viral, bacterial, fungal and

CC parasitic infections. The present sequence is a cDNA encoding a protein  
CC of the invention.

XX Sequence 1187 BP, 374 A, 252 C, 271 G, 288 T, 2 other;

Query Match 1.0%; Score 19; DB 21; Length 1187;

Best Local Similarity 100.0%; Pred. No. 41;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 370 AGCTGTTCTTCTGAATCTT 388

DB 354 AGCTGTTCTTCTGAATCTT 372

RESULT 25

AA158527

ID AA158527 standard; cDNA; 1240 BP.

XX AA158527;

XX 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 730.

XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
XX peripheral nervous system; neuropathy; central nervous system; CNS;  
XX Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;  
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemocarcin;  
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
XX leukemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0483725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0593042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
XX Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;  
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

XX P-PSDB; AAM39371.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
XX such as central nervous system injuries -

XX Claim 1; SEQ ID NO 730; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and  
XX the encoded polypeptides (AAM38642-AAM42213) with neurotropic,  
XX immunosuppressant and cytostatic activity. The polynucleotides are useful  
XX in gene therapy. A composition containing a polypeptide or polynucleotide  
XX of the invention may be used to treat diseases of the peripheral nervous  
XX system, such as peripheral nervous injuries, peripheral neuropathy and  
XX localized neuropathies and central nervous system diseases, such as  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
XX utilisation of the activities such as: Immune system suppression,  
XX Activin/inhibin activity, chemocarcin/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukemias and  
CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

XX Sequence 1240 BP, 351 A, 272 C, 306 G, 311 T, 0 other;

Query Match 1.0%; Score 19; DB 22; Length 1240;

Best Local Similarity 100.0%; Pred. No. 42;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 370 AGCTGTTCTTCTGAATCTT 388

DB 420 AGCTGTTCTTCTGAATCTT 438

RESULT 26

AAH33729

ID AAH33729 standard; cDNA; 1311 BP.

XX AAH33729;

XX 03-SEP-2001 (first entry)

DE Human colon cancer antigen encoding cDNA SEQ ID NO:785.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
XX colorectal carcinoma; ss.

XX Homo sapiens.

XX WO200122920-A2.

XX 05-APR-2001.

PF 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

XX (HDMA-) HDMA GENOME SCT INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI: 2001-235357/24.

XX P-PSDB; AAG74298.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX useful for preventing, diagnosing and/or treating colorectal cancers -

XX Claim 1; Page 2755; 9803bp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
XX cancer-associated nucleic acid molecules (N) and proteins (P), where  
XX the proteins are collectively known as colon cancer antigens. The colon  
XX cancer antigens have cytostatic activity and can be used in gene  
XX therapy and vaccine production. N and P may be used in the prevention,  
XX diagnosis and treatment of diseases associated with inappropriate P  
XX expression. For example, N and P may be used to treat disorders  
XX associated with decreased expression by rectifying mutations or deletions  
XX in a patient's genome that affect the activity of P by expressing  
XX inactive proteins or to supplement the patient's own production of P.  
XX Additionally, N may be used to produce the colon cancer-associated Ps,  
XX by inserting the nucleic acids into a host cell and culturing the cell  
XX to express the proteins. N and P can be used in the prevention, diagnosis  
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
XX and AAG77789 represent sequences used in the exemplification of the  
XX present invention.  
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
XX missing at time of publication, meaning no sequences are present for  
XX SEQ ID NO:1027 to 1052, 7921 and 7922.

SQ Sequence 1311 BP; 349 A; 282 C; 316 G; 361 T; 3 other;  
Query Match 1.0%; Score 19; DB 22; Length 1311;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 370 AGCTGTTCTTCTGAATCTT 388  
Db 387 AGCTGTTCTTCTGAATCTT 405

RESULT 27  
AAI60313/c  
ID AAI60313 standard; cDNA; 1684 BP.  
XX AC AAI60313;  
XX DT 22-OCT-2001 (first entry).  
XX DE Human polynucleotide SEQ ID NO 4302.  
XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX OS Homo sapiens.  
XX PN WO200153312-A1.  
XX PD 26-JUL-2001.  
XX PF 26-DEC-2000; 2000WO-US34263.  
XX PR 21-JAN-2000; 2000US-0488725.  
XX PR 25-APR-2000; 2000US-0552317.  
XX PR 09-JUL-2000; 2000US-0598042.  
XX PR 19-JUL-2000; 2000US-0620312.  
XX PR 03-AUG-2000; 2000US-0653450.  
XX PR 14-SEP-2000; 2000US-0662191.  
XX PR 19-OCT-2000; 2000US-0693036.  
XX PR 29-NOV-2000; 2000US-0727344.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX WPI; 2001-442253/47.  
XX P-PSDB; AAM41157.  
XX PT Novel nucleic acids and polypeptides, useful for treating disorders  
XX PT such as central nervous system injuries -  
XX PS Claim 1; SEQ ID NO 4302; 10078pp; English.  
XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
XX CC the encoded polypeptides (AAM38642-AAAM42213) with nootropic,  
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide  
XX CC of the invention may be used to treat diseases of the peripheral nervous  
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and  
XX CC localised neuropathies and central nervous system diseases, such as  
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
XX CC utilisation of the activities such as: Immune system suppression,  
XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and  
XX CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX SQ Sequence 1684 BP; 485 A; 388 C; 358 G; 453 T; 0 other;  
Query Match 1.0%; Score 19; DB 22; Length 1684;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 370 AGCTGTTCTTCTGAATCTT 388  
Db 1346 AGCTGTTCTTCTGAATCTT 1328

RESULT 28  
AAH68588  
ID AAH68588 standard; cDNA; 1684 BP.  
XX AC AAH68588;  
XX DT 13-SEP-2001 (first entry)  
XX DE Human protein HP10522 coding sequence.  
XX KW Human; gene therapy; tumour; ss.  
XX OS Homo sapiens.  
XX PN WO200142302-A1.  
XX PD 14-JUN-2001.  
XX PF 06-DEC-2000; 2000WO-JP08631.  
XX PR 06-DEC-1999; 99JP-0346863.  
XX PR 08-DEC-1999; 99JP-0346864.  
XX PR 08-FEB-2000; 2000JP-0031062.  
XX PR 10-FEB-2000; 2000JP-0034090.  
XX PR 10-FEB-2000; 2000JP-0034091.  
XX PR 14-FEB-2000; 2000JP-0035829.  
XX PR 14-FEB-2000; 2000JP-0035899.  
XX PR 14-MAR-2000; 2000JP-0071161.  
XX PR 30-MAY-2000; 2000JP-0160851.  
XX PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
XX PI Kato S, Eguchi C, Saeki M;  
XX WPI; 2001-381646/40.  
XX P-PSDB; AAG93303.  
XX PT Human protein originated from tumor cell line, applicable as drug,  
XX PT reagent for studying intracellular protein networks and protein source  
XX PT for drug screening, also encoded cDNA for gene diagnosis and gene  
XX PT therapy -  
XX PS Claim 3; Pages 349-353; 471pp; Japanese.  
XX CC The present sequence is a human protein coding sequence. The human  
XX CC protein, preferably originated from tumor cell line, is applicable as a  
XX CC drug, a reagent for studying intracellular protein networks and a protein  
XX CC source for screening proteins for binding low molecular weight drugs. The  
XX CC human protein coding sequence is useful for gene diagnosis and gene  
XX CC therapy, expression vectors and transformant cells for detection of  
XX CC ligands and receptors.  
XX SQ Sequence 1684 BP; 444 A; 360 C; 390 G; 490 T; 0 other;  
Query Match 1.0%; Score 19; DB 22; Length 1684;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 370 AGCTGTTCTTCTGAATCTT 388

DB 351 AGCTGTTCTTCGATCTT 369

RESULT 29

AAH16819

ID AAH16819 standard; cDNA; 1685 BP.

AAH16819;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:16081.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0115126.

29-JUL-1999; 99JP-0243036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0113776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Claim 8; SEQ ID 16081; 2537bp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 1685 BP; 442 A; 362 C; 392 G; 489 T; 0 other;

Query Match 1.0%; Score 19; DB 22; Length 1685;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 370 AGCTGTTCTTCGATCTT 388  
DB 350 AGCTGTTCTTCGATCTT 368

RESULT 30

AAH16007

ID AAH16007 standard; cDNA; 1690 BP.

AAH16007;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:14646.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Claim 8; SEQ ID 14646; 2537bp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 1690 BP; 444 A; 361 C; 390 G; 495 T; 0 other;

Query Match 1.0%; Score 19; DB 22; Length 1690;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Qy 370 AGCTGTTCTTCTGAATCTT 388  
 Db 357 AGCTGTTCTTCTGAATCTT 375

RESULT 31  
 AAD24382  
 ID AAD24382 standard; cDNA; 1749 BP.  
 XX  
 AC AAD24382;  
 XX  
 DT 07-MAR-2002 (first entry)  
 XX  
 DE Human RNA metabolism protein-9 (RMEP-9) cDNA.  
 XX  
 KW Human; RNA metabolism protein-9; RMEP-9; gout; nervous system disorder;  
 KW autoimmune; inflammatory; cell proliferative; developmental; thyroiditis;  
 KW gene therapy; epilepsy; dementia; stroke; Alzheimer's disease; amnesia;  
 KW Parkinson's disease; prion disease; insomnia; endocrine disorder; AIDS;  
 KW Acquired Immune Deficiency Syndrome; mental disorder; allergy; anaemia;  
 KW asthma; atherosclerosis; Crohn's disease; rheumatoid arthritis; vaccine;  
 KW glomerulonephritis; multiple sclerosis; diabetes mellitus; osteoporosis;  
 KW cancer; cirrhosis; hepatitis; psoriasis; transgenic animal; antiulcer;  
 KW tranquiliser; drug screening; pancreatitis; renal tubular acidosis;  
 KW systemic lupus erythematosus; colitis; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 48..1046  
 FT /\*tag= a  
 FT /product= "Human RMEP-9 protein"  
 FT sig\_peptide 48..137  
 FT /\*tag= b  
 FT mat\_peptide 138..1043  
 FT /\*tag= c  
 FT /product= "Mature human RMEP-9 protein"

WO200183524-A2.  
 08-NOV-2001.  
 27-APR-2001; 2001WO-US13862.  
 28-APR-2000; 2000US-200184P.  
 04-MAY-2000; 2000US-201875P.  
 04-MAY-2000; 2000US-202090P.  
 06-JUN-2000; 2000US-210232P.  
 25-JUL-2000; 2000US-220553P.  
 (INCY-) INCYTE GENOMICS INC.  
 Lal P, Yue H, Tang YT, Lu DAM, Azimzai Y, Au-Young J, Hillman JL;  
 Baughn MR, Yao MG, Burford N, Batra S, Policky JJ;  
 WPI; 2002-034502/04.  
 P-PSDB; AAE15246.  
 New human RNA metabolism protein for diagnosing or treating nervous  
 system disorders, autoimmune/inflammatory disorders, cell proliferative  
 disorders and developmental disorders -  
 Claim 11; Page 175; 196pp; English.  
 The invention relates to human RNA metabolism proteins (RMEP) and their  
 corresponding cDNA molecules. RMEP and its DNA are used for diagnosing,  
 treating and preventing nervous system disorders (epilepsy, dementia,  
 stroke, Alzheimer's disease, Huntington's disease, Parkinson's disease);  
 prion diseases; fatal familial insomnia, nutritional and metabolic  
 diseases of the nervous system; inherited, metabolic, endocrine and  
 toxic myopathy; mental disorders (mood, anxiety, schizophrenic disorders)  
 amnesia and Tourette's disorder; autoimmune/inflammatory disorders (AIDS-

CC acquired immune deficiency syndrome, allergies, anaemia, asthma, gout,  
 CC atherosclerosis, Crohn's disease, diabetes mellitus, glomerulonephritis,  
 CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid  
 CC arthritis, osteoporosis, pancreatitis, systemic lupus erythematosus,  
 CC ulcerative colitis, and infections); cell proliferative disorders (cancer  
 CC arteriosclerosis, cirrhosis, hepatitis, psoriasis); and developmental  
 CC disorders (renal tubular acidosis). RMEP DNA is useful in drug screening.  
 CC techniques, gene therapy and for creating transgenic animals. The present  
 CC sequence is human RMEP-9 cDNA.  
 XX

SQ Sequence 1749 BP; 474 A; 373 C; 396 G; 506 T; 0 other;

Query Match 1.0%; Score 19; DB 24; Length 1749;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 AGCTGTTCTTCTGAATCTT 388

Db 386 AGCTGTTCTTCTGAATCTT 404

RESULT 32

AAV99969/c

ID AAV99969 standard; cDNA; 2364 BP.

XX AC AAV99969;

XX DT 07-JUN-1999 (first entry)

XX DE Nucleotide sequence of the human smoothened gene.

XX KW Human; smoothened gene; smoothened peptides; SPS; liver repair;  
 KW cell proliferation; cell differentiation; spermatogenesis;  
 KW osteogenesis; chondrogenesis; apoptosis; degenerative disorder;  
 KW neoplastic disorder; hyperplastic disorder; varicella-zoster;  
 KW tumour; ds.  
 XX

OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 1..2364  
 FT /\*tag= a  
 FT /product= "smoothened polypeptide"

XX PN WO9901468-A2.

XX PD 14-JAN-1999.

XX PF 01-JUL-1998; 98WO-US13793.

XX PR 21-JUL-1997; 97US-0897798.

XX PR 01-JUL-1997; 97US-0051407.

XX PA (INGH/) INGHAM P W.

XX PA (ONTO-) ONTOGENY INC.

XX PA (VHEU/) VAN DEN HEUVEL M.

XX DR WPI; 1999-105996/09.

XX DR P-PSDB; AAW81064.

XX FT New isolated smoothened genes - used to develop products for  
 FT treating e.g. cancers, neurodegenerative disorders, nervous system  
 FT injury, immunological diseases or infections

XX PS Claim 12; Pages 90-91; 107pp; English.

XX CC This is the nucleotide sequence of the novel human smoothened  
 CC gene used in the method of the invention to produce smoothened  
 CC polypeptides (SPs). The SPs can modulate at least one of  
 CC proliferation, differentiation or survival of a cell which  
 CC expresses the SP. The products can be used to modulate  
 CC spermatogenesis, osteogenesis, chondrogenesis or neuronal cell  
 CC differentiation or to enhance survival of neuronal cells, e.g. to

XX

FT /product= smc

/product= Smoothened protein

PI Stone DM, De Sauvage FJ, Rosenthal A;  
XX  
XX WFI; 2001-006229/01.  
DR P-PDB; AAB28646.

XX  
XX Antibodies to vertebrae smoothed proteins which relate to Hedgehog  
PT and patched signaling molecules involved in cell proliferation and  
PT differentiation -  
XX  
XX Example 3; Fig 4; 38pp; English.

XX The present sequence encodes human Smoothed. Smoothened interacts with  
CC Hedgehog and Patched signalling molecules involved in cell proliferation  
CC and differentiation. The rat Smoothed cDNA was isolated from an  
CC embryonic day 9-10 rat CDNA library. The rat Smoothed cDNA was then  
CC used as a probe to isolate the human homologue from a human embryonic  
CC lung cDNA library. Antibodies which specifically bind to the vertebrate  
CC smoothened polypeptides were identified.

XX Sequence 2972 BP; 525 A; 929 C; 918 G; 600 T; 0 other;

SQ

Query Match            1.0%; Score 19; DB 22; Length 2972;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1280 TTGGCAATCATCTTGCTCT 1298  
   ||||| |  
Db 1724 TTGGCAATCATCTTGCTCT 1706

RESULT 36  
ABL32627  
ID ABL32627 standard; DNA; 19082 BP.  
  
AC ABL32627;  
XX  
XX 26-MAR-2002 (first entry)  
DE Human immune system associated gene SEQ ID NO: 600.  
  
KW Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antidiabetic; antianemic; cytostatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW rheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX WO200200928-A2.  
PN XX  
XX PD 03-JAN-2002.  
PP XX  
PP 02-JUL-2001; 2001WO-EPO7537.  
PR XX  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX WFI; 2002-130909/17.  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
XX Claim 1; SEQ ID NO 600; 32pp + Sequence Listing; German.  
PS  
XX The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.

SQ Sequence 19082 BP; 5624 A; 429 C; 4025 G; 9004 T; 0 other;

Query Match 1.0%; Score 19; DB 24; Length 19082;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 509 CGAGAGCAATTTTATTG 527  
 DB 11533 CGAGAGCAATTTTATTG 11551

RESULT 37

ABL09700 ABL09700 standard; cDNA; 22660 BP.

AC ABL09700;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 23582.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191537P.

PR 11-JUL-2000; 2000US-0611150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB65597.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

PS Claim 1; SEQ ID NO 23582; 21bp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1030 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB857737-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 22660 BP; 6703 A; 4609 C; 4758 G; 6590 T; 0 other;

Query Match 1.0%; Score 19; DB 23; Length 22660;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 AAAAACAATGTACACAAG 284  
 DB 19690 AAAAACAATGTACACAAG 19708

RESULT 38  
 AAL39687/c  
 ID AAL39687 standard; DNA; 31352 BP.

AC AAL39687;

DT 05-SEP-2002 (first entry)

DE Genomic DNA encoding the human smoothened Drosophila homologue (SMO).

KW Cytostatic; polymorphic variant; single nucleotide polymorphism; SMOH;

KW human smoothened Drosophila homologue; basal cell carcinoma; BCC;

KW gene therapy; antisense gene therapy; gene; ds.

OS Homo sapiens.

PH Key Location/Qualifiers

FT CDS 4044..27353

FT /product= "Protein of human smoothened Drosophila

FT homologue gene"

FT /note= "This coding region contains 11 introns"

FT 4044..4374

FT /tag= b

FT /number= 1

FT 4375..18279

FT /tag= c

FT /number= 1

FT 18280..18485

FT /tag= d

FT /number= 2

FT 18486..20104

FT /tag= e

FT /number= 2

FT 20105..20314

FT /tag= f

FT /number= 3

FT 20315..20511

FT /tag= g

FT /number= 3

FT 20512..20684

FT /tag= h

FT /number= 4

FT 20685..21051

FT /tag= i

FT /number= 4

FT 21052..21271

FT /tag= j

FT /number= 5

FT 21272..21365

FT /tag= k

FT /number= 5

FT 21366..21489

FT /tag= l

FT /number= 6

FT 21490..23660

FT /tag= m

FT /number= 6

FT 23661..23753

FT /tag= n

FT /number= 7

FT 23754..24190

FT /tag= o

FT /number= 7

FT 24191..24299

FT /tag= p

FT /number= 8

```
FT intron 24300..25264
FT /tag= q
FT /number= 8
FT exon 25265..25450
FT /tag= r
FT intron 25451..25866
FT /tag= s
FT exon 25867..26015
FT /tag= t
FT intron 26016..26537
FT /tag= u
FT exon 26538..26672
FT /tag= v
FT intron 26673..26925
FT /tag= w
FT exon 26926..27353
FT /tag= x
FT variation 27354..27498
FT /tag= y
FT variation 27499..27643
FT /tag= z
FT variation 27644..27788
FT /tag= aa
FT variation 27789..27933
FT /tag= ab
FT variation 27934..28078
FT /tag= ac
FT variation 28079..28223
FT /tag= ad
FT variation 28224..28368
FT /tag= ae
FT variation 28369..28513
FT /tag= af
FT variation 28514..28658
FT /tag= ag
FT variation 28659..28803
FT /tag= ah
FT variation 28804..28948
FT /tag= ai
FT variation 28949..29093
FT /tag= aj
FT variation 29094..29238
FT /tag= ak
FT variation 29239..29383
FT /tag= al
FT variation 29384..29528
FT /tag= am
FT variation 29529..29673
FT /tag= an
FT variation 29674..29818
FT /tag= ao
FT variation 29819..29963
FT /tag= ap
FT variation 29964..30108
FT /tag= aq
FT variation 30109..30253
FT /tag= ar
FT variation 30254..30398
FT /tag= as
FT variation 30399..30543
FT /tag= at
FT variation 30544..30688
FT /tag= au
FT variation 30689..30833
FT /tag= av
FT variation 30834..30978
FT /tag= aw
```

```
FT variation /tag= aw
FT replace(26928,T)
FT /tag= ax
FT variation replace(27041,A)
FT /tag= ay
FT variation replace(27248,A)
FT /tag= az
FT variation replace(27442,T)
FT /tag= ba
FT variation replace(27572,A)
FT /tag= bb
XX WO200229004-A2.
XX 11-APR-2002.
XX PD
XX 04-OCT-2001; 2001WO-US31304.
XX PF
XX 04-OCT-2000; 2000US-237871P.
XX PR
XX (GENA-) GENAISSANCE PHARM INC.
XX PA
XX Bentivegna SC, Choi JY, Koshy B, Lee HH, Sausker EA;
XX PI
XX WPI; 2002-519113/55.
XX DR
XX P-PSDB; AAO21722.
XX
XX New genetic variants of smoothened Drosophila homolog (SMOH) gene
XX useful for therapeutic purposes and for expressing SMOH protein useful
XX in identifying drugs to treat basal cell carcinomas -
XX
XX Claim 1; Fig 1; 179pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising a sequence
XX which is a polymorphic variant of a reference sequence for the human
XX smoothened Drosophila homologue (SMOH) gene or its fragment, or a
XX polymorphic variant of a reference sequence for a SMOH cDNA or its
XX fragment. A new isolated polypeptide is useful for screening for drugs
XX targeting the polypeptide. A new method is useful for identifying an
XX association between a trait such as a clinical response to a drug
XX targeting SMOH and a haplotype or haplotype pair of SMOH gene. The
XX methods have applicability in developing diagnostic tests and therapeutic
XX treatments for basal cell carcinomas (BCCs). The isolated polynucleotide
XX is useful for studying the expression and function of SMOH and expressing
XX SMOH protein for use in screening for candidate drugs to treat diseases
XX related to SMOH activity. The polymorphism and haplotype data are useful
XX for validating whether SMOH is a suitable target for drugs to treat BCCs,
XX screening for the drugs and reducing bias in clinical trials of the
XX drugs. The isolated polynucleotide is useful for therapeutic purposes.
XX The new method, an oligonucleotide and kit of the invention are useful
XX for determining whether an individual has one of the haplotypes or the
XX haplotype pairs. The polynucleotides of the invention can be used to
XX treat disorders by gene therapy and antisense gene therapy. This
XX polynucleotide sequence represents the genomic DNA encoding the human
XX smoothened Drosophila homologue (SMOH) of the invention.
XX
XX Sequence 31352 BP; 7528 A; 7628 C; 7963 G; 8233 T; 0 other;
XX
XX Query Match 1.0%; Score 19; DB 24; Length 31352;
XX Best Local Similarity 100.0%; Pred. No. 48;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1280 TTGGCAATCATCTTGCTCT 1298
XX |||||
XX Db 25926 TTGGCAATCATCTTGCTCT 25908
XX
XX RESULT 39
XX ABN26694
XX ID ABN26694 standard; cDNA; 235 BP.
XX XX
XX AC ABN26694;
XX XX
```

DT 24-JUN-2002 (first entry)  
 XX  
 XX Human ORFX polynucleotide sequence SEQ ID NO:21865.  
 DE  
 XX  
 XX Human: open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KM hyperproliferative disorder; psoriasis; benign tumor; haemorrhage;  
 KM degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KM cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KM hypertension; hypothyroidism; cholesterol ester storage disease;  
 KM immune deficiency; immune disorder; infectious disease;  
 KM autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KM myasthenia gravis; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200192523-A2.  
 XX  
 PD 06-DEC-2001.  
 XX  
 PF 29-MAY-2001; 2001WO-US10836.  
 XX  
 PR 30-MAY-2000; 2000US-206132P.  
 XX  
 PR 29-AUG-2000; 2000US-228716P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shinkets RA, Leach MD;  
 XX  
 PI WPI; 2002-106308/14.  
 DR P-PSDB; ABP10942.  
 XX  
 XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders -  
 XX  
 PS Disclosure; SEQ ID 21865; 1037pp; English.  
 XX  
 XX The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP0010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 235 BP; 64 A; 42 C; 62 G; 67 T; 0 other;  
 QY  
 Query Match 1.0%; Score 18; DB 24; Length 235;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 709 CATGACTAAATTTTCTT 726  
 120 CATGACTAAATTTTCTT 137

RESULT 40  
 AA191598/C  
 ID AA191598 standard; CDNA; 357 BP.  
 XX  
 XX  
 AC AA191598;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 XX Human polynucleotide SEQ ID NO 11658.  
 DE  
 XX  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KM nervous system disorders; arthritis; inflammation; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200164835-A2.  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 26-FEB-2001; 2001WO-US04927.  
 XX  
 PR 28-FEB-2000; 2000US-0515126.  
 XX  
 PR 18-MAY-2000; 2000US-0577409.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 PI WPI; 2001-514838/56.  
 DR P-PSDB; AA011667.  
 XX  
 XX Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 XX  
 PS Claim 1; SEQ ID NO 11658; 1399pp + Sequence Listing; English.  
 XX  
 XX The invention relates to human polynucleotides (AA179941-AA193841) and  
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibit activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 357 BP; 115 A; 61 C; 88 G; 92 T; 1 other;  
 QY  
 Query Match 1.0%; Score 18; DB 22; Length 357;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 110 AGCCCTATTTTAAACGTC 127  
 257 AGCCCTATTTTAAACGTC 240  
 RESULT 41  
 ABN90719  
 ID ABN90719 standard; DNA; 378 BP.  
 XX  
 AC ABN90719;  
 XX  
 DT 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:182.  
DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX antibacterial; gene therapy; gene; ds.  
KW Staphylococcus epidermidis.  
XX US6380370-B1.  
PN 30-APR-2002.  
XX 13-AUG-1998; 98US-0134001.  
XX 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX Doucette-Stamm LA, Bush D;  
PI WPI; 2002-381255/41.  
XX P-PSDB; ABP38174.  
DR Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
FT Disclosure; SEQ ID 182; 267pp; English.  
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP5124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. the sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX SQ Sequence 378 BP; 150 A; 57 C; 65 G; 106 T; 0 other;  
  
Query Match 1.0%; Score 18; DB 24; Length 378;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 266 AAAAAACAATGTACACAA 283  
Db ||||||||||||||||  
211 AAAAAACAATGTACACAA 228  
  
RESULT 42  
ABK44466/C  
ID ABK44466 standard; cDNA; 459 BP.  
XX AC ABK44466;  
XX 05-JUN-2002 (first entry)  
XX cDNA encoding colon tumour protein, SEQ ID No 17.  
XX Human; colon tumour; vaccine; colon cancer; immunogenic;  
KW immunotherapy; gene; ss.  
XX Homo sapiens.  
XX WO200212328-A2.  
PN 14-FEB-2002.  
XX 31-JUL-2001; 2001WO-US24219.  
XX

PR 03-AUG-2000; 2000US-223283P.  
PR 28-MAR-2001; 2001US-279763P.  
PR 29-JUN-2001; 2001US-302051P.  
XX (CORI-) CORIXA CORP.  
XX King GE, Meagher MJ, Xu J, Secrist H;  
XX WPI; 2002-241739/29.  
XX New colon cancer polypeptides and polynucleotides, useful as vaccines,  
PT for diagnosing, preventing, and treating colon cancer, and as markers  
FT for the progression of cancer -  
XX Claim 1; SEQ ID No 17; 147pp; English.  
XX The invention relates to polynucleotides encoding colon tumour proteins.  
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical  
CC compositions, such as vaccines, for the diagnosis, prevention, and  
CC treatment of colon cancer. Polynucleotide sequences may be used as  
CC hybridisation probes or primers, and in the design and preparation of  
CC ribozyme molecules for inhibiting expression of tumour polypeptides and  
CC proteins in tumour cells. The compositions are useful for stimulating an  
CC immune response against cancer, particularly for the immunotherapy of  
CC colon cancer, and as markers for the progression of cancer.  
CC ABK44450-ABK46237 represent coding sequences of human colon tumour  
CC proteins of the invention.  
CC Note: With the exception of SEQ ID No 1 and 2, the sequence data  
CC for this patent did not form part of the printed specification but was  
CC supplied by the European Patent Office.  
XX SQ Sequence 459 BP; 112 A; 93 C; 107 G; 147 T; 0 other;  
  
Query Match 1.0%; Score 18; DB 24; Length 459;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 887 TCCGTTTGGGAATGGAT 904  
Db ||||||||||||||||  
36 TCCGTTTGGGAATGGAT 19  
  
RESULT 43  
ABL93238/C  
ID ABL93238 standard; cDNA; 478 BP.  
XX AC ABL93238;  
XX 10-JUN-2002 (first entry)  
XX Arabidopsis thaliana nucleic acid sequence Ref:2027003 SEQ ID NO:3.  
DE Arabidopsis thaliana; insecticide; fungicide; plant; mapping; diagnosis;  
KW genetic modification; gene; ss.  
XX Arabidopsis thaliana.  
XX US2002023280-A1.  
PN 21-FEB-2002.  
XX 26-JAN-2001; 2001US-0770444.  
XX 27-JAN-2000; 2000US-178502P.  
XX (GORL/) GORLACH J.  
PA (ANY/) AN Y.  
PA (HAMI/) HAMILTON C M.  
PA (PRIC/) PRICE J L.  
PA (RAIN/) RAINES T M.  
PA (YUY/) YU Y.  
PA (RAME/) RAMEAKA J G.  
PA (PAGE/) PAGE A.

PA (MATH/) MATHEW A. V.  
 PA (LEDF/) LEDFORD B. L.  
 PA (WOES/) WOESSNER J. P.  
 PA (HAAS/) HAAS W. D.  
 PA (GARC/) GARCIA C. A.  
 PA (KRICK/) KRICKER M.  
 PA (SLAT/) SLATER T.  
 PA (DAVI/) DAVIS K. R.  
 PA (ALLE/) ALLEN K.  
 PA (HOFF/) HOFFMAN N.  
 PA (HURB/) HURBAN P.  
 XX  
 PI Gorlach JG, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;  
 PI Ramaek JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD,  
 PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;  
 PI Hurban P;  
 XX  
 DR WPI, 2002-267486/31.  
 XX  
 PT New Arabidopsis thaliana nucleic acid, for identifying homologous  
 PT genes, producing compositions that modulate the expression or function  
 PT of its encoded protein, and mapping functional regions of a protein -  
 XX  
 PS Claim 1, SEQ ID 3; 44pp; English.  
 XX  
 CC The present invention describes an Arabidopsis thaliana nucleic acid (I)  
 CC comprising a sequence capable of hybridising under stringent conditions  
 CC to a sequence (S1) selected from any one of the 999 sequences given in  
 CC AB193236 to AB194334. (I) have insecticide and fungicide activities, and  
 CC they can be used as protein expression modulators. (I) can be used in  
 CC identifying homologous or related genes, in producing compositions that  
 CC modulate the expression or function of their encoded proteins, mapping  
 CC functional regions of the proteins, and in studying associated  
 CC physiological pathways. (I) can also be used: (1) for the genetic  
 CC manipulation of cells, particularly plant cells; (2) in screening assays  
 CC of various plant strains to determine the strains that are best capable  
 CC of withstanding a particular disease or environmental stress; (3) for  
 CC enhancing or inhibiting production of a biosynthetic product in a plant;  
 CC (4) as probes in mapping and in diagnosis, in genetic modification and  
 CC for screening purposes, to generate additional copies of the nucleic  
 CC acids, to generate ribozymes or as triple-strand forming oligonucleotides;  
 CC single-stranded DNA probes or as triple-strand forming oligonucleotides;  
 CC and (5) for generating genetically modified transgenic organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.  
 XX  
 SQ Sequence 478 BP; 131 A; 93 C; 116 G; 138 T; 0 other;  
 XX  
 QY Query Match 1.0%; Score 18; DB 24; Length 478;  
 Best Local Similarity 103.0%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 DB 835 TGGAGGTGTTTCTTTT 852  
 |||||||  
 97 TGGAGGTGTTTCTTTT 80  
 XX  
 RESULT 44  
 AB1919019  
 ID AB1919019 standard; cDNA; 576 BP.  
 XX  
 AC AB1919019;  
 XX  
 DT 13-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 19010.  
 XX  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KM pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX

PN W0200160860-A2.  
 XX  
 XX 23-AUG-2001.  
 PD  
 XX  
 PF 20-FEB-2001; 2001WO-US05171.  
 XX  
 PR 17-FEB-2000; 2000US-183119P.  
 PR 16-MAR-2000; 2000US-189862P.  
 PR 25-MAY-2000; 2000US-207454P.  
 PR 09-JUN-2000; 2000US-213114P.  
 PR 18-JUL-2000; 2000US-219007P.  
 PR 13-DEC-2000; 2000US-255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Endege WO, Monahan JE;  
 PI  
 XX  
 DR WPI, 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer -  
 XX  
 PS Claim 1, Page 3121; 11750pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;  
 CC (g) determining whether prostate cancer has metastasized in a patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
 XX  
 SQ Sequence 576 BP; 189 A; 89 C; 104 G; 174 T; 20 other;  
 XX  
 QY Query Match 1.0%; Score 18; DB 23; Length 576;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 DB 895 GGGATGGATTTAAGG 912  
 |||||||  
 418 GGGATGGATTTAAGG 435  
 XX  
 RESULT 45  
 AB066275  
 ID AB066275 standard; DNA; 613 BP.  
 XX  
 AC AB066275;  
 XX  
 DT 21-AUG-2002 (first entry)  
 XX  
 DE Arabidopsis thaliana polynucleotide SEQ ID NO 852.  
 XX  
 KW Arabidopsis thaliana; thale cress; plant; transgenic; GMO; disease;  
 KM stress; metabolic pathway; biosynthetic pathway; nutrition; fungicide;  
 XX insecticide; antibiotic; ds.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN US2002059663-A1.  
 XX  
 PD 16-MAY-2002.  
 XX  
 PF 26-JAN-2001; 2001US-0770149.



```

XX 27-JAN-2000; 2000US-178506P.
XX
XX (GORL/) GORLACH J.
PA (ANY/) AN Y.
PA (HAMI/) HAMILTON C M.
PA (PRIC/) PRICE J L.
PA (RAIN/) RAINES T M.
PA (YUY/) YU Y.
PA (RAME/) RAMEAKA J G.
PA (PAGE/) PAGE A.
PA (MATH/) MATHAW A V.
PA (LEDF/) LEDFORD B L.
PA (WOES/) WOESSNER J P.
PA (HAAS/) HAAS W D.
PA (GARC/) GARCIA C A.
PA (KRIC/) KRICKER M.
PA (SLAT/) SLATER T.
PA (DAVI/) DAVIS K R.
PA (ALLE/) ALLEN K.
PA (HOFF/) HOFFMAN N.
PA (HURB/) HURBAN P.
XX
XX Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
PI Hurban P;
XX
XX WPI; 2002-479224/51.
XX
XX New nucleic acid that hybridizes to Arabidopsis thaliana sequences,
PT useful e.g. for preparing transgenic plants with increased resistance
PT or altered metabolism
XX
XX Claim 1; SEQ ID NO 852; 40pp + Sequence Listing; English.
XX
XX The invention relates to nucleic acids (I) that hybridise under stringent
CC conditions to any of 999 sequences (ABQ65424-ABQ66422) or their
CC fragments. (I) are used to express the corresponding polypeptides (II) or
CC to produce genetically modified plant cells or transgenic plants, which
CC may have improved resistance to disease or stress, or altered
CC metabolic/biosynthetic pathways (for production of commercial,
CC nutritional or medicinal products), or generally any trait of interest,
CC or can be used to screen for biologically active agents (e.g. fungicides,
CC insecticides and antibiotics).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO at seqdata.uspto.gov/sequence.html?DocID=999909770149.
XX
XX SQ Sequence 613 BP; 177 A; 146 C; 115 G; 175 T; 0 other;

```

```

Query Match      1.0%; Score 18; DB 24; Length 613;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 517 ATTTTATTGCTCAAG 534
   |||||
Db 73 ATTTTATTGCTCAAG 90

```

```

Search completed: December 17, 2002, 09:05:10
Job time : 405 secs

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 17, 2002, 12:31:36 ; Search time 34.5 Seconds  
(without alignments)

10388.091 Million cell updates/sec

Title: US-09-868-987-1  
Perfect score: 1312  
Sequence: 1 atggacttcgcattgtgct.....atgacaattcagataatgc 1864

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV=xlh  
-MODE=frame-n2p.model -DEV=xlh  
-Q=/cgn2\_1/USFTO\_spool/US09868987/runat 16122002 083007 23561/app query.fasta\_1.2055  
-DB=PIR\_73 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=45 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09868987 @CGN 1.1 32 @runat 16122002 083007 23561 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR\_73 : \*  
1: pir1 : \*  
2: pir2 : \*  
3: pir3 : \*  
4: pir4 : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2874	86.8	1402	2 H86560	protein export [im
2	2874	86.8	1402	2 E72062	protein export pro
3	2083	62.9	1400	2 A71514	probable protein e
4	2068	62.4	1400	2 A81672	secDF protein, pro
5	840	25.4	857	2 A2768	protein-export mem
6	840	25.4	876	2 B97549	probable protein-e
7	715.5	21.6	845	2 A53337	protein translocas
8	685	20.7	759	2 D89946	protein-export mem
9	678.5	20.5	571	2 H69704	protein-export mem
10	668.5	20.2	768	2 F75350	probable protein-e
11	653.5	19.7	775	2 A73386	protein translocas
12	630	19.0	795	2 T34673	probable SecDF pro
13	620.5	18.7	754	2 A11627	protein-export mem
14	618.5	18.7	754	2 AG1265	protein-export mem

15	522.5	15.8	873	2 C90594	protein-export mem
16	443	13.4	472	1 S76266	hypothetical prote
17	436	13.2	474	2 A11821	protein-export mem
18	431	13.0	503	1 F64713	protein-export mem
19	429	13.0	526	2 D71805	protein-export mem
20	427	12.9	501	1 C70384	protein-export mem
21	426	12.9	583	1 H71326	probable protein-e
22	424	12.8	620	2 F83169	secretion protein
23	408.5	12.3	616	1 I64056	secretion protein
24	403.5	12.0	528	2 H81162	protein-export mem
25	398.5	12.0	518	2 E71663	protein-export mem
26	395.5	11.9	518	2 F97811	protein-export mem
27	391	11.8	417	2 H97180	preprotein translo
28	390	11.8	532	2 B87496	protein-export mem
29	390	11.8	323	2 G81312	protein-export mem
30	384.5	11.6	322	2 AD0387	protein-export mem
31	383.5	11.6	615	2 AE0387	secretion protein
32	376	11.4	617	2 B82285	protein-export mem
33	369.5	11.2	306	2 E83169	protein-export mem
34	368	11.1	612	2 F82429	protein-export mem
35	365.5	11.0	437	2 D83804	protein-export mem
36	365	11.0	323	2 AH0552	protein-export mem
37	364	11.0	323	1 J00697	preprotein translo
38	364	11.0	323	2 D80686	protein-export mem
39	364	11.0	323	2 H85536	protein secretion,
40	364	11.0	618	2 C81926	probable protein-e
41	364	11.0	618	2 H81178	protein-export mem
42	363.5	11.0	315	2 C82285	protein-export mem
43	359.5	10.9	615	1 H64769	preprotein translo
44	359.5	10.9	615	2 C90686	protein-export mem
45	359.5	10.9	615	2 G85536	protein-export mem

## ALIGNMENTS

### RESULT 1

H86560  
protein export [imported] - Chlamydothila pneumoniae (strain J138)  
C:Species: Chlamydothila pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: H86560  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Ouchi, K.; Shiba, T.; Ishi  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349; PMID:10871362  
A:Accession: H86560  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1402 -STO-  
A:Cross-references: GB:BA000008; NID:g8978935; PIDN:BA098770.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: secD\_secF

Alignment Scores:	6.03e-244	Length:	1402
Pred. No.:	2874.00	Matches:	583
Score:	99.49%	Conservative:	0
Percent Similarity:	99.49%	Mismatches:	3
Best Local Similarity:	86.78%	Indels:	3
Query Match:	2	Gaps:	0
DB:			

US-09-868-987-1 (1-1864) x H86560 (1-1402)

QY	2	TCGACTTCGCATATTGTTCAGAGGGATACCGGCTGCTAATGACAAATATTCGCA	61
Db	817	TpThrSerAlaTy-CysGlnGluGlyIleSerGlyThrAlaAsnGlyGlnTySerAla	836
QY	62	AACCTGGATGGCGATGCTCTAGTATTGACGTTATATGTCAGCAGCCCTATTTTA	121
Db	837	AsnArgGlyTrpArgMetAlaValIleAspGlyTyMetValSerSerProIleLeu	856
QY	122	AACGTCCTTGAATAATCATCCAGTGTCTCAGGAAATTTTACCACCGTCAGTACGAGC	181

```

Db 857 AsnValPheLeuLysAsnHisLeuValSerGlyLysPheThrHisArgGluValSer 876
Qy 182 AAACCTGCGCTCAGATTAAATCTGAGCGCATGCTTTGTTGCCGAGGTTCTCAGTAA 241
Db 877 LysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheValProGluValLeuSerGlu 896
Qy 242 GAGAGCATCTCTTCGATCTTGGGAAAAAACAATGACACAGGATTAATCTGACATGC 301
Db 897 GluThrIleSerSerAspLeuGlyLysLysGlnCysThrGlnGlyIleIleSerAlaCys 916
Qy 302 TGTCGCTTGGCAATGCTTAATGTTTATGATGAGCTATTAATGATTTGAGCGGTCATC 361
Db 917 CysGlyLeuAlaMetLeuIleValLeuMetSerValIlyrTrpPheGlyGlyValIle 936
Qy 362 GCTTCGGGAGCTGTTCTTCGAATCTTTGCTTATCTGGACGCTCAGATTTGAT 421
Db 937 AlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrpAlaAlaLeuGlnIlyrLeuAsp 956
Qy 422 GCGGCATCACTTGTGAGAGCTCGCGGAGTGTCTTGTCTATGGGATGGCCGTAGAT 481
Db 957 AlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaValAsp 976
Qy 482 GCAAATGTTCTTGTATTCGAAGAATCCGAGAGAATTTTATTTGCTCAAACTCTTAA 541
Db 977 AlaAsnValLeuValPheGluArgIleArgGluGluPheLeuLeuSerGlnSerLeuLys 996
Qy 542 AAATCGTGAAGAAAGATTAACAGAGCTTTTGAGCCATTTTGATTTGATTAATGACT 601
Db 997 LysSerValGluLysGlyIlyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeuThr 1016
Qy 602 AACAATATGGCTCAGACACTTCTTCTCTCAGATACAGGCGCATTAAGAGGTTTCT 661
Db 1017 ThrValLeuAlaSerAlaLeuLeuPheLeuAspThrGlyProIleLysGlyPheAla 1036
Qy 662 TTGACATTCATTTTAGGAATTTTCTCTCAATGTTTACGCTCTTTTCATGATTAATTT 721
Db 1037 LeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLysPhe 1056
Qy 722 TTCTTCATGCTGTGATTAATTAAGACCAACATACAGCTTGATATGATTAATGATTC 781
Db 1057 PhePheMetLeuTrpMetAsnLysThrGlnHisTrpGlnLeuHisMetMetAsnLysPhe 1076
Qy 782 GTGGGAGATTAAGCATGATTTCTTGAGAGATGCAAAAACTTTGGCTGTTCTTGAAGT 841
Db 1077 ValGlyIleLysHisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGlySer 1096
Qy 842 GTTTTCTTTAGGTTGCTGCTCGGCTGCTGGGTTTGAGAGCTTGGAATTCGTTTGGGATG 901
Db 1097 ValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValIleuGlyMet 1116
Qy 902 GATTTTAAAGAGGATATGCTTTACCTTAATCAAAAAGCATGTCATCAGCATGTT 961
Db 1117 AspPheLysGlyGlyIlyrAlaPheThrPheAsnProLysGlnHisGlyIleSerAspVal 1136
Qy 962 GCTCAAAATGCGTGCAAAAGTTGTCATTAACATGAGAGAGCTGCTTTCTTGAAGAC 1021
Db 1137 AlaGlnMetArgGlyLysValValHisLysLeuGlnGlnIleGlyLeuSerSerArgAsp 1156
Qy 1022 TTCGATTAACAAATTTGATCTTTCAGAAAAGATCAAAATCAATTTTGTGATAAGCT 1081
Db 1157 PheArgIleGlnThrPheGlySerSerGluLysIleLysIleLysIleLysSerAspLysAla 1176
Qy 1082 TTAAGCATATCAAGCA-3ATACAGAGCTCTCTCTCAAAATTAACGATCATGAGCTGCG 1140
Db 1177 LeuSerIlyrThrLysAlaAspThrSerLeuSerProLysIleAsnAspHisGluValAla 1196
Qy 1141 TTATTT-GTGGGATTTGTTGAGAAACAGGCTGATGTTTCTTAACGAAACTTAACGAA 1199
Db 1199 LeuAlaValGlyLeuLeuSerGlnThrGlyLeuAspPheSerThrGlnThrLeuAsnGln 1216
Qy 1216 CAAAA-TTTTGTCAAGTAAAGTAAAGCAAACTTTCGAAAGAAATGCGTATCAGGCG 1258

```

---

```

Db 1217 ThrGlnAsnPheTrpSerLysValSerSerLysLeuSerLysMetArgTyGlnAla 1236
Qy 1259 ACCATGCGGCTTTTAGAGCTTTTGCAATCATCTTGCTATGAGTTCGCTTTGAA 1318
Db 1237 ThrIleGlyLeuLeuGlyAlaLeuAlaIleIleLeuLeuIlyrValSerLeuArgPheGlu 1256
Qy 1319 TGGCAATATGCTTTCAATGCGGCTATGCGCTTAAATTCATGACCTTTGGCTTACCTGCA 1378
Db 1257 TrpGlnIlyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaTrpCysAla 1276
Qy 1379 GCTTGTATTAAGCAGATTTCTTTTGAAGAATAATCAATTAATTTGCAAGCATTTGGT 1438
Db 1277 ValLeuPheIleAlaHisPhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGly 1296
Qy 1439 GCTTAAATGACTGATTTGGGATTTTCATTAACAAATATCTTTCATTTTGTGATG 1498
Db 1297 AlaLeuMetThrValLeuGlyIlyrSerLeuAsnAsnThrLeuIleIlePheAspArgIle 1316
Qy 1499 CGTGAAGATGCGCAAGGAAACCTGTACCCCTTACATGATTTTATGATATGCGCTT 1558
Db 1317 ArgGluAspArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeu 1336
Qy 1559 CAAAAGCGTTACGCCGCAACGGTAATGACACAGCTCAACTCTTCACTTTTGTATATG 1618
Db 1337 GlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMet 1356
Qy 1619 CTTTGTATTAATGCGGCTCTCTGCTTTAATTTTGATTTATTAATGATGAGATT 1678
Db 1357 LeuLeuPheIleGlyLysSerValPheAsnPheAlaPheIleMetThrIleGlyLe 1376
Qy 1679 CTTTGAAGACTTTATCGCTCTTATATATGACACACCGCTGTTTATATGTCCTG 1738
Db 1377 LeuLeuGlyIlyrThrLeuSerSerLeuIlyrIleAlaProPheLeuLeuPheMetValArg 1396
Qy 1739 AAAAGAAATCGCTCAAAA 1756
Db 1397 LysGluAsnArgSerLys 1402

```

**RESULT 2**  
 E72062  
 protein export proteins secD/secE (fusion) - Chlamydomonas pneumoniae (strains CWL029  
 C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
 R:Accession: E72062, C81605  
 R:Author: S. Mitchell, W. Marathe, R. Lammell, C. Fan, J. Olinger, L. Grimwood, J  
 Nature Genet. 21, 385-389, 1999  
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
 A:Reference number: A72000; MUID:99206606; PMID:10192388  
 A:Accession: E72062  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1402 <RNA>  
 A:Cross-references: GB:AE001641; GB:AE001363; NID:g4376655; PIDN:AAD18704.1; PID:g4376  
 A:Experimental source: strain CWL029  
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidberg, J.F.; White, O.; Hickey  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis MOpn and Chlamydia pneumoniae AR39  
 A:Reference number: AB1500; MUID:20150255; PMID:10684935  
 A:Accession: C81605  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1402 <RNA>  
 A:Cross-references: GB:AE002179; GB:AE002161; NID:g7189108; PIDN:AAF38059.1; PID:g7189  
 A:Experimental source: strain AR39, HL cells  
 C:Genetics:  
 A:Gene: secD/secE; CP018

**Alignment Scores:**  
 Pred. No.: 6.03e-244 Length: 1402  
 Score: 2874.00 Matches: 583  
 Percent Similarity: 99.49% Conservative: 0  
 Best Local Similarity: 99.49% Mismatches: 3